the newsletter of the International Computer Science Institute

The ICSI GAZETTE

volume twelve | issue two | fall 2014

featured research: privacy

In June 2013, the *Guardian* published its first report based on documents provided by former National Security Agency contractor Edward Snowden. Just more than a year later, privacy seems to be on everyone's mind. In a survey conducted in January by Truste, a data management company that prioritizes consumer safety and privacy, 74 percent of users said they were more worried about online privacy than they had been a year earlier. Pew Research found that half of Americans were worried about their personal information on the Internet in 2013, up from a third in 2009. A quick Google search for "online privacy" calls up articles with ominous headlines such as "Online Privacy is Dead." But what exactly is online privacy, and why is it important?

"It depends on who you ask and what they know," says Blanca Gordo, an artificial intelligence researcher. Among other things, she's been working with the Teaching Privacy team, which has built a web site that explains what happens to personal information when it goes online, along with a set of apps to demonstrate this. The team comprises researchers from three ICSI groups and from UC Berkeley. "This is the kind of work you can't do alone. It's complex, but that doesn't mean it can't be articulated."

A Pedagogy of Privacy

Teaching Privacy was first funded through a supplementary grant to the Geo-Tube Project, which is led by Gerald Friedland, director of Audio and Multimedia research, and Robin Sommer of Networking and Security. In this project, researchers are studying the ways in which an attacker can aggregate public and seemingly innocuous information from different media and web sites to harm users. The project seeks to help users, particularly younger ones, understand the risks of sharing information online and the control – or lack thereof – they have over it.

Friedland and Sommer began studying this topic in 2010, when they coined the word "cybercasing" to describe the use of geo-tagged information available online to mount attacks in the real world. In a paper presented at the USENIX Workshop on Hot Topic in Security in August 2010, they showed that it was easy to discover where many photos and videos posted online were taken by extracting metadata with highly precise location information, known as geo-tags, that are embedded by many higher-end digital cameras and smart phones. They then cross-referenced these with other information such as the text accompanying a photo in a Craigslist ad to find users who might be away from the house during the day or who had gone on an extended vacation. In one search of YouTube videos, for example, the researchers were able to find users with homes near downtown Berkeley by searching the embedded geo-location data. They then searched for videos posted by the same users that had been filmed over 1,000 miles away. Within fifteen minutes, the researchers found a resident of Albany, California who was vacationing in the Caribbean, along with a dozen other users who might be vulnerable to burglary.

Friedland says that the interest sparked by this research motivated him to seek funding for the Geo-Tube project. "In 2010, people weren't aware that this could be done," he said. "The issue was to simply raise awareness."

Much of Audio and Multimedia's work over the last four years has been to show what can be inferred about

continued on page 5

in this issue:

new director:
deborah crawford

On August 15, Dr. Deborah L. Crawford joined ICSI as its new director. Dr. Crawford was formerly the senior vice provost of research at Drexel University, where she was instrumental in the creation of its College of Computing and Informatics.

Crawford received her PhD from the University of Bradford, England, and worked at AT&T Bell Labs, UC Santa Barbara, and the Jet Propulsion Laboratory before joining the National Science Foundation in 1993. During her 17-year tenure at NSF, she held a number of executive positions. In 2002, she was named deputy assistant director of the Directorate for Computer and Information Science and Engineering, where she played a leadership role in the development of programs that encouraged the computer science community to imagine bold new futures for the field, such as the Expeditions in Computing program. She left NSF to join Drexel in 2010.

Crawford has twice received the Presidential Rank Award, given by the president to recognize senior federal executives for leadership, strength, integrity, industry, and commitment to excellence in public service.

“Deb’s unflagging commitment to service and scientific leadership and her remarkable record of creatively and resolutely solving problems make her uniquely suited to lead ICSI,” said Dr. Prabhakar Raghavan, the chair of ICSI’s Board of Trustees. “We are thrilled she has decided to join us.”

“I am delighted to be joining the ICSI community,“ said Crawford. “While I’ve been able to work with my new colleagues as a member of the ICSI Board of Trustees, I will now work with this world-class community of researchers on a daily basis, united in our passion and commitment to advance computing in ways that are currently just gleams in our eyes. What a privilege it is to join such a distinguished and dynamic group.”

She succeeds Professor Nelson Morgan as director.

taking the long view

by Deborah Crawford, Director

I am delighted to have this opportunity to pen my first remarks as director, having joined ICSI just a handful of weeks ago. In a career that has spanned almost three decades, I've had the great fortune of working with really smart people on inspiring projects in computing. With each passing year, I've only come to enjoy more the rewards that stem from seeing new ideas come to life enabled by the creative contributions and insights of computing's best and brightest. Needless to say, I look forward to continuing this fine tradition at ICSI.

ICSI is a place where great minds from around the world come together to engage with significant opportunities and challenges in computing — a place where open inquiry and creative thinking are the norm, and where new ideas and disruptive innovations have a real opportunity to change the world as we know it. With research groups working on the most challenging scientific and socio-technical computing problems of our time — from well-established emphases in networking and security, artificial intelligence, speech and vision, to emerging areas of strength in multimedia, data science and analytics, brain science and health IT – ICSI and its researchers-in-residence host visiting researchers from all corners of the globe, from the junior undergraduate researcher to the most senior faculty fellow. It is at ICSI that inspiring research visions and ambitious innovation goals crystallize, enabled by the experiences that proximity to Bay area excellence in computing discovery and innovation provides.

ICSI thrives on great partnerships, partnerships that include individuals and institutions in the academic community, both here and abroad, companies large and small, foreign and domestic, public schools, federal, state and local governments, and other nations and international bodies. These collaborations work best when we and our partners bring complementary capabilities and expertise to the table, so that together we are able to do what we cannot accomplish alone.

In the coming months, we, the ICSI community, will develop a strategic plan to help guide the future of our institute. We will share and discuss our hopes and dreams for computing, and in the process synthesize a ten-year vision for ICSI and corresponding strategic directions and actions required to deliver on it.

As we plan for this undertaking, we take time today to highlight some of the great work currently underway here. This newsletter describes the work of several ICSI researchers and research projects focused on addressing the pressing socio-technical challenges of securing cyberspace. These challenges are likely to continue to pose a limitless number of research opportunities, for as the 2014 report of the Computer Science and Telecommunications Board of the National Academies notes:

> "Cybersecurity problems result from the inherent nature of information technology, the complexity of information technology, and human fallibility in making judgements about what actions and information are safe or unsafe from a cybersecurity perspective, especially when such actions and information are highly complex. None of these factors is likely to change in the foreseeable future, and thus there are no silver bullets – or even combinations of silver bullets – that can "solve the problem" permanently.

Cybersecurity problems result from the inherent nature of information technology, the complexity of information technology, and human fallibility in making judgements about what actions and information are safe or unsafe from a cybersecurity perspective, especially when such actions and information are highly complex. None of these factors is likely to change in the foreseeable future, and thus there are no silver bullets — or even combinations of silver bullets — that can "solve the problem" permanently.

Furthermore, they are only likely to be addressed by the creative contributions of researchers with expertise in computing and in adjacent and complementary fields. I hope you enjoy learning more about our interdisciplinary work aimed at creating a secure and trustworthy cyberspace.

Please know that as we, the ICSI community, mobilize to chart our path into the future, we will stay true to our past — computing the future in partnership with colleagues the world over is in our DNA. We remain committed to contributing advances in computing that rock the world, delight our partners, and make our researchers, from the most junior to the quite senior, incredibly proud of what they do.

I hope you will take the opportunity to join us in our strategic planning activities. I believe that some of our best years lie ahead, and that together we can work to ensure that ICSI remains the vibrant and inspiring place it is today.

featured alum:
massimo maresca

Professor Massimo Maresca of the University of Genoa, an ICSI alum as well as the head of the Scientific Office of the Italian Consulate in San Francisco, is supervising a new project at ICSI related to end-user computing, big data, and the Internet of things.

Professor Maresca first visited ICSI's Realization Group, which developed massively parallel systems, as a postdoc in the early 1990s. He soon switched his attention to distributed networks, an area of research led at ICSI by Professor Domenico Ferrari in the Networking Group. Professor Maresca went on to work at the University of Genoa, first as an assistant professor and later as an associate professor, while continuing to visit ICSI. In 1995, he moved to the University of Padua, maintaining close ties with collaborators in Genoa. These collaborations led to the foundation of the Research Center on Computer Platform Engineering (CIPI).

The center's research focuses on platforms – standard sets of functionalities for performing services. Specifically, the center identifies and characterizes platforms, and researches how to connect them so that they can cooperate to perform complex tasks.

The project at ICSI is on the service composition paradigm for distributed applications. This paradigm is a way of visualizing and organizing the relationships among services that, together, make up a larger service composed by simple basic functionalities (i.e., web APIs) belonging to different domains such as social networks, e-commerce, news, and government data. Researchers from the center will visit ICSI to study how this paradigm can be applied to two areas: the Internet of Things (IoT), a network in which most physical objects we interact with on a daily basis will one day be connected, and Big Data. Michele Stecca, a CIPI researcher, will spend his postdoctoral fellowship working at ICSI on different components of the project.

For CIPI's work on the IoT, each real object within this network is represented in the network as an abstract representation, or an agent. This agent corresponds to a small, less complex service in a composite service network. In this way, the internet of things is visualized as a composite service. The problem is how to orchestrate these agents.

A past project sponsored by the European Commission, for example, applied these ideas to smart cars, considering different functions and pieces of the car – the brakes, the gas, etc. – as agents, or atomic services and linking them together in a network, or composite service.

While at ICSI, Stecca will be working in the other domain, Big Data. The center has launched a start-up, SpreadsheetSpace, that links spreadsheets together in a network, creating a larger composite spreadsheet.

"When you receive data analyses, you sometimes want to change something to understand the data," said Professor Maresca. By linking the data sources, SpreadsheetSpace allows a regular end-user to do that. This is a part of end-user computing. "You don't have to be a programmer to use Excel. Most people are able to do something in Excel," he said. "We address these people." When applied to the business domain, "our potential is huge because the world uses Excel."

In Italy, Professor Maresca is an avid sailor, spending two weeks each summer off the coast of Cinque Terre in the Italian Riviera.

featured research, continued

continued from page 1

users from relatively little information. In ongoing research, for example, they show how it is increasingly possible to estimate where videos were shot even if they are not geo-tagged, by looking at text data such as titles and tags, visual cues such as textures and colors, and sounds such as bird songs and ambulance sirens. In other work, they've found that comparing the audio tracks of different videos can show whether they were uploaded by the same person – meaning that online accounts can be linked even if they are created under different names.

But four years and dozens of released NSA documents later, Friedland says, the issue is no longer just informing people that there are privacy risks, but also suggesting mitigations.

"Our expertise is to answer the questions, What can you do technically to protect privacy?" he said. If, for example, you take a photograph of yourself but you don't want anyone to be able to tell where you are or whom you're with: "How much do you need to change your face or the background to maintain privacy? You interview Snowden and you blur the background; can you use background noise to identify the city he's in?"

The Teaching Privacy team works to give students the knowledge and tools to protect their privacy online. The effort, which will continue on as Teaching Resources for Online Privacy Education (TROPE) led by Friedland and Networking and Security researcher Serge Egelman, has developed 10 principles for social media privacy, with technical explanations written for a lay audience and real-life examples. The team has also built several apps that show users exactly what their social media accounts are sharing. The Ready or Not? app, released in August 2013, takes GPS data embedded in Twitter and Instagram posts to create a heat map of where users have been posting from and when.

The Ready or Not? app

The team comprises not only computer scientists, who design the apps and ground the principles in technical research, but also educators and social scientists like Gordo. In separate work, she's been evaluating the California Connects program, which sought to spread broadband adoption among disconnected populations. Her and other artificial intelligence researchers' role in Teaching Privacy, she says, is to ground the work in an understanding of the ways people think.

"There is no pedagogy for digital functioning," she said. "This is why I'm interested in building a language to teach technology."

One difficulty facing those who want to spread Internet adoption – or, in the case of Teaching Privacy, educate young users about how privacy works – is that few data exist showing how new users understand technology. "We have a responsibility to ground a baseline of understanding by paying attention to what they're thinking, how they're processing information," she said. "We have to understand how they think."

"A Very Nuanced Issue"

Serge Egelman, who joined Networking and Security in 2013, has also been working on human factors in privacy. In one project, he is investigating how certain individual differences (e.g., personality traits) correlate with privacy preferences and security behaviors; the goal is to build systems that automatically adapt to user preferences. For example, an operating system could analyze behaviors on the computer and on the web, learn a user's privacy preferences, and then automatically change website privacy preferences based on those inferred preferences.

He points to default privacy settings on web sites like Facebook, which are the same for all people who sign up at the same time. "Obviously, everyone doesn't have the same privacy preferences," he said. "So, can we look at other behaviors to infer your default preferences?"

continued on next

featured research, continued

continued from previous

Of course, these individualized preferences would require human intervention – you would need to tweak who sees posts you're tagged in, for example – but they would be a "vast improvement over the default."

He works with cognitive psychologist Eyal Pe'er of Bar-Ilan University, Israel, to conduct online surveys that measure classic psychometrics, such as extraversion, neuroticism, and risk aversion. They also ask about privacy and security behaviors. In some surveys, participants are asked about increasingly sensitive information, a technique borrowed from behavioral economics. The point at which participants begin refusing to answer gives some insight into how open they are.

"There's a spectrum of privacy preferences," he said. "It's a very nuanced issue."

In addition to his position at ICSI, Egelman is a researcher at UC Berkeley's Electrical Engineer and Computer Sciences Department. There, he is working on a project that investigates how third-party mobile apps use data, with the ultimate goal of giving users the ability to make more fine-grained decisions. He and his colleagues are investigating privacy alerts for apps that chart a middle course between two prevailing practices, one that provides too much information and one that provides too little. On Android phones, for example, every time users download an app, the phone provides a list of the personal data that the app may collect, choice of discontinuing the installation if they do not agree to granting the application access. This is given without context, however. Take GPS, for example. How does an app use the phone's GPS data? Is it for a legitimate and helpful location-based service, or is it simply to monetize the data? With this approach, the user becomes habituated to accepting the warnings; users click through the list automatically. "It's the boy-who-cried-wolf syndrome," he said.

On the other hand, apps for the Apple iOS platform provide very little information, and because the approval process for iOS apps is so opaque, it's difficult to understand how collected data are used at all.

He and his colleagues are investigating what aspects of privacy users actually care about and comparing that with situations where those aspects are implicated. The goal is to infer when data use will concern users and then only show privacy warnings during those situations. "We've done experiments, and we've found that given a practice, some people view this as an encroachment on privacy, while others view it as a desirable feature of an app."

The Paradox of the Internet

Istemi Ekin Akkus, a fourth-year PhD student at the Max Planck Institute for Software Systems in Kaiserslautern, Germany, who visited Networking and Security this spring and summer, points out that there is a paradox at the heart of the web: the Internet economy relies upon the trust of customers and also upon data about them that they may not want revealed. Analytics, or information about how users browse the web and use mobile apps, provides companies with important customer data that enable ads targeted at customer interests. The advertising industry says these ads are much more effective than traditional advertising, and Akkus says their high revenue allows web sites that display them to provide content for free.

Akkus is interested in developing technologies that both prevent the tracking of users and sustain the online economy. At ICSI, Akkus worked on the Priv3 Firefox extension, which was initially released in 2011. The extension stops certain social media sites from knowing when you've visited a page. You've probably noticed, while browsing the web, the social widgets that let you post an article to Facebook or Twitter – and perhaps failed to notice, since it's done invisibly, that those sites are tracking your activity if you're logged in. This is done through cookies, small text files that sites store on your web browser. Say you log into Twitter and then go to a news site that allows you to tweet the article directly from the page. Your browser downloads the social widget from Twitter's web site. While doing so, it also sends the cookie Twitter has stored on

The Internet economy relies upon the trust of customers and also upon data about them that they may not want revealed.

your computer when you logged in to Twitter. Upon receiving this cookie, Twitter knows you've visited the page even if you choose not to tweet the article. Priv3 stops Twitter, along with Facebook, LinkedIn, and Google+, from receiving such cookies right away: it waits until you interact with the widget (by liking the article on Facebook, for example, or sharing it on your LinkedIn profile). In this way, the user gets to enjoy the benefits of social widgets as well as protection from third parties silently tracking browsing.

While useful, Priv3 depends on a blacklist of specific social media sites, which must be maintained as new social media sites gain popularity. In addition, cookies are used by more than just social media sites. In 2012, the Berkeley Center for Law and Technology found that all of the most popular 100 web sites used cookies, more than two-thirds of which came from third parties. These cookies follow users from site to site, allowing companies to keep track of their interests (at least, as inferred from their web browsing history) and place ads that are most relevant to potential customers. This may end up revealing more than users want to share. Akkus, with other members of the group, are working to extend Priv3 to all third-party tracking without the need of a blacklist. With this generalization, the tracking by these third parties using cookies will be mostly ineffective: although they will be able to set new cookie values, they will not be able to receive them unless the user wants them to.

> "Give me that data and I will have a blackmail target in every senate office."
>
> *- Nicholas Weaver on Verizon records collected by the NSA*

Nicholas Weaver, a senior researcher in Networking and Security and a self-described paranoid, disputes that the Internet economy necessarily depends on the collection of browsing data. "There's tons of proof that you can do advertising without using tracking and still make money at it," he said, pointing to television and print ads as examples. "They don't actually need to know that I visited site A when I visit site B. All they need to know is that I'm the kind of person who would visit site B."

"In some ways, it's a failure of the Internet ecosystem that it relies on such aggressive tracking because of the pushback," he said. For example, many people use ad-blocking technologies, like AdBlock, not because they dislike the ads but because they are worried about tracking.

Government Surveillance

Weaver says ads, particularly ad banks used by smartphone apps, present another concern: according to documents leaked by Snowden, the U.S. government "piggybacks" on the tracking information sent by ads, such as the phone's location. This information can be correlated with, say, an Instagram account that includes the user's name.

Weaver has been focusing some of his paranoia on the Snowden revelations. He sees three major concerns with the scope and depth of NSA surveillance. First, it carries the potential for abuse. "Look at what J. Edgar Hoover did. What would Hoover have been able to accomplish with the NSA's level of surveillance?" he said.

The first Snowden disclosure, on June 5, 2013, showed that the NSA had been collecting the phone records of Verizon customers, regardless of whether they were suspected of wrongdoing. "It is vastly dangerous to have that data in raw form," said Weaver. "Give me that data and I will have a blackmail target in every senate office."

Fortunately, he says, the NSA has used this data "remarkably" responsibly. But the revelations raise a second concern: many techniques used by the NSA can be done with small, off-the-shelf hardware.

"This is what really worries me, that the replicability is so easy that you can do it almost anywhere," he said. "The only limitation on adversaries is vantage point."

continued on next

featured research, continued

continued from previous

Russia already uses a system similar to one NSA program revealed by Snowden; in 2013, the U.S. State Department warned travelers to the Winter Olympics in Sochi about surveillance, recommending that "essential devices should have all personal identifying information and sensitive files removed or 'sanitized.'" Weaver also said the Great Firewall of China, which is now used to censor the country's Internet access, could easily be used for surveillance.

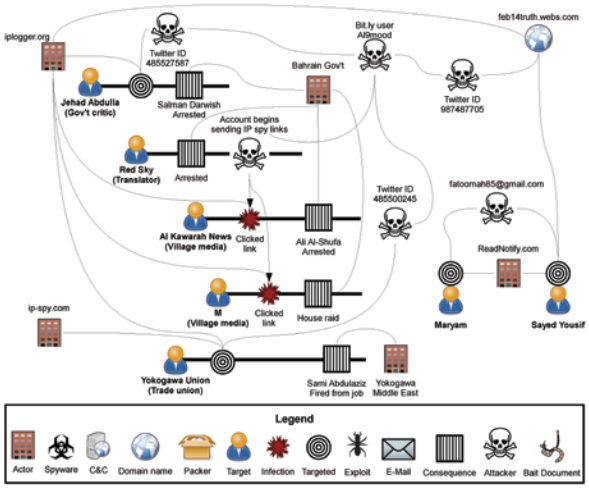
Not only is this level of surveillance easy, but it now carries the imprimatur of the U.S. Weaver pointed to reports last year that the agency had spied on German Chancellor Angela Merkel's cell phone in addition to other members of the European Union. "The NSA has given the permission to do this," he said. "This acts as an open invitation."

But even if the technology is not abused either by the U.S. or by foreign adversaries, Weaver says, the knowledge that this information is being collected has a chilling effect on Internet use. "It affects us through self-censorship," he said. "If there is a possibility, however remote, that a disclosure may cause you damage, and there's a worry about that, it changes your behavior, often to your detriment."

Friedland, the Audio and Multimedia director, says the chilling effect, and the potential for abuse, are particularly problematic in a free country. "That's why secrets are a fundamental part of our democracy," he said. "There needs to be a point where you can switch off the record."

Technology to Repress

Bill Marczak, a PhD student at UC Berkeley and a researcher in Networking and Security, has been studying how governments in what are not considered free countries use technology to repress activists, journalists and other critics of their governments. Marczak, Networking and Security Director Vern Paxson, and members of the Citizen Lab at the University of Toronto's Munk School of Global Affairs have been found that the governments in Bahrain, Syria, and the United Arab Emirates use malware to identify and attack activists. A paper describing the research was presented at the USENIX Security Symposium in August.



The ecosystem of Bahrain's IP Spy attacks

Marczak wants to bring international media attention to this practice and develop technical solutions and training to give activists the ability to protect themselves against cyber-attacks from regimes.

"I bring a certain type of research which is so often missing from activism," he said. He noted that activist research often lacks rigorous evidentiary standards.

Marczak's interest in the Middle East dates back to the three and a half years he spent in Bahrain as a high school student. "When you go to high school in a certain place, you get to know people there," he said. "It really becomes part of who you are."

During the Arab Spring uprisings, which came to a head in Bahrain on February 14, 2011, the country's government began increasing its use of physical weapons to fight protesters and turned to Western PR firms to burnish its reputation abroad. At the time, Marczak was interested in cloud computing and programming languages; he decided to switch his focus to political surveillance and electronic attacks.

He got to know several activists working in the Middle East on Twitter; in February 2012, they founded Bahrain Watch. The organization, which is independent and works on research and advocacy, operates "without any sort of structure, as an Internet collective." Marczak says this helps them defend against attacks from repressive regimes: "Their minds are calibrated to look for activists as people who are organizing."

Marczak also works with several human rights groups, including Privacy International, a charity based in London, and the Citizen Lab.

In April and May 2012, several activists interested in Bahrain were sent suspicious attachments, which, when Bill analyzed them, turned out to contain spyware developed by Gamma International, a company based in Germany. The spyware, FinFisher, can record passwords, log keystrokes, and take screenshots of the infected computer. Marczak and his colleagues traced the spyware's command and control server to an address in Bahrain.

In order to identify other activity, they sent probes to servers and watched their behavior. They then scanned the Internet, searching for similar behavior. In addition to the server in Bahrain, they've found similar spyware associated with servers in six other countries considered to be ruled by oppressive regimes.

Marczak says the Citizen Lab in Toronto has been a particularly strong partner in identifying spyware. The lab, which primarily focuses on government surveillance and censorship systems, has a global network with connections to people in countries around the world.

At the USENIX Security Symposium in August, Marczak, Paxson, and Marczak's colleagues at Citizen Lab and UC Los Angeles, presented a paper about their research into FinFisher, as well as the use of Hacking Team's Remote Control System in the United Arab Emirates. Like Gamma, Hacking Team markets its software exclusively to governments. In Syria, they investigated the use of off-the-shelf remote access trojans. They found that the attacks were probably a factor in the year-long imprisonment of one activist and the publication of embarrassing videos of another, who was subsequently discredited.

The paper describes the "careful social engineering" used by attackers – in Bahrain, for example, several activists received email messages with attachments that were claimed to be reports of torture or pictures of jailed citizens. The paper also notes that this is the first step in a broader rigorous study of attacks target at individuals by governments.

Marczak said, "It's difficult because researchers have little visibility into what activists are doing. I suspect more's going on than we can see. Part of the goal is to gain better visibility and engagement with activists."

What's to be Done

Even without the risk of abuse on the level seen by Marczak and his colleagues, online privacy remains an important issue for most users of the Internet. Gordo says, "In different fields and different discourses, you're finding this question: What is privacy?"

Gordo says Europeans and Americans differ in their approach to privacy. In Europe, privacy is a human right; "in the United States, it's a contract. But if it's a contract, people need to know what they're signing on to."

Egelman points out that Europeans and Americans also differ in how they view the government: Europeans believe that the government serves to protect them from corporations; Americans tend to worry more about the intrusion of government into the private sector.

He says, "It ultimately comes down to exerting control over your information: preventing its dissemination, controlling its dissemination, or simply being aware of how your information is disseminated." For his work, in which he tries to make security warnings and privacy defaults responsive to individual users, "there is no universal truth. It's about informed consent."

Gordo says, "For me, it's urgent because policy is being formulated, and the end user isn't being taken into account."

And after informing users and policy-makers, researchers will have a role in developing mitigations. Egelman is interested in the future of wearable computing – think Google Glass. "The approach we should be taking is, what are the issues we can imagine when these devices are pervasive, and what are the things we can do to mitigate those issues," he said. "We should be thinking about these issues now. We can prevent issues down the road when the devices are pervasive."

Friedland said, "Independent institutions like ICSI will play a role" in the work to secure privacy "They will go in and say, 'Look, this is what we can do.'"

Teaching Privacy's 10 Principles and its apps, Egelman's work on inferred privacy preferences, Akkus's work on Priv3, and several other projects are all part of that work. And you can always just heed Weaver's simple advice:

"Encrypt everything."

news briefs


Vern Paxson

Professor **Vern Paxson**, who directs Networking and Security research, has been named a recipient of the 2015 IEEE Internet Award, given annually to recognize exceptional contributions to the advancement of Internet technology for network architecture, mobility, or end-use applications. Paxson and KC Claffy of the Cooperative Association for Internet Data Analysis in San Diego, California, were recognized for their contributions to the field of Internet measurement, including security and network data analysis, and for distinguished leadership in and service to the Internet community by providing open-access data and tools. The Internet Award is sponsored by Nokia.

Chief Scientist **Scott Shenker** and Professor David Culler, a member of ICSI's board of trustees, are co-authors of a paper that has received a Test-of-Time Award at the Networked Systems Design and Implementation Symposium (NSDI). Their paper, "Trickle: A Self-Regulating Algorithm for Code Propagation and Maintenance in Wireless Sensor Networks," won the best paper award when it appeared in 2004 and was also written by Philip Levis and Neil Patel. Culler is also a co-author of a second paper being honored, "Operating System Support for Planetary-Scale Network Services." This is the first year NSDI has given the Test-of-Time Award, which was presented during the symposium April 2-4 in Seattle, Washington.

Professor **Krste Asanovic**, the former director of architecture research, has received the 2014 Diane S. McEntyre Award for Excellence in Teaching from UC Berkeley's Computer Science Division. The award, given annually to a member of the computer science faculty, recognizes commitment to and respect for students. Asanović received his PhD from UC Berkeley in 1998 while working as a student in the Realization Group (what is now the Speech Group) and founded the Architecture Group ten years later.


Scott Shenker

Professor **Scott Shenker**, ICSI's chief scientist and the director of Research Initiatives, has won the 2013-2014 Jim and Donna Gray Faculty Award for Excellence in Undergraduate Teaching from UC Berkeley's Computer Science Division. The award is given annually to faculty who demonstrate excellence in undergraduate teaching. In the past ten years, five members, alumni, and affiliates of ICSI have won the award, including former architecture research director Krste Asanović in 2009-2010 and Networking and Security research director Vern Paxson in 2010-2011.


Wolfgang Wahlster

Professor **Wolfgang Wahlster**, a member of ICSI's Executive Committee, was inducted into the Hall of Fame of the most important German-speaking information technology personalities of the past 40 years. The induction ceremony was hosted by the magazine Computerwoche, the German version of IDG's Computerworld, at the end of June in Berlin. He was honored for his pioneering research on human-machine interaction. Other inductees include his colleagues Karlheinz Brandenburg, the director of the Fraunhofer Institute of Digital Media Technology and the inventor of the MP3 audio compression format; Peter Grünberg of Forschungszentrum Jülich, who won the Noble Prize for discovering giant magnetoresistance; and Niklaus Wirth of ETH Zurich, a Turing Award recipient and the designer of the Pascal programming language. Professor Wahlster is the director and CEO of the German Research Center for Artificial Intelligence (DFKI GmbH) and a professor of computer science at Saarland University.

students and visiting scholars

Since its inception, ICSI has had a strong international program consisting primarily of ties with specific countries. Formal agreements exist with Germany, Italy, and Singapore. In addition, we have visitors associated with specific research and projects.

AI
Björn Fritsche
Elisabeth Wehling (Germany)
Alexander Ziem (Germany)

Audio and Multimedia
Damian Borth (Germany)
Xiao-Yong Wei

Speech
Van Tung Pham (Singapore)

Research Initiatives
Itamar Afek
Shachar Afek
Yaron Anavi
Yael Baran
Jacob Goldberger
Hayit Greenspan
Regev Schweiger
Michele Stecca (Italy)

Networking
Istemi Ekin Akkus
Ignacio de Castro Arribas
Haixin Duan
Zakir Durumeric
Roya Ensafi
Dirk Hasselbalch
Asim Jamshed
Jian Jiang
Ben Jones
Christof Leng (Germany)
Jinjin Liang
Antonio Nappa
Philipp Richter
Luigi Rizzo
Amin Tootoonchian
David Wang
Yifei Xu

Vision
Jiashi Feng
Stefanie Jegelka
Damian Mrowca
Claudia Nieuwenhuis (Germany)
Marcus Rohrbach (Germany)


Elisabeth Wehling


Van Tung Pham


Roya Ensafi


Michele Stecca

100 million images and videos

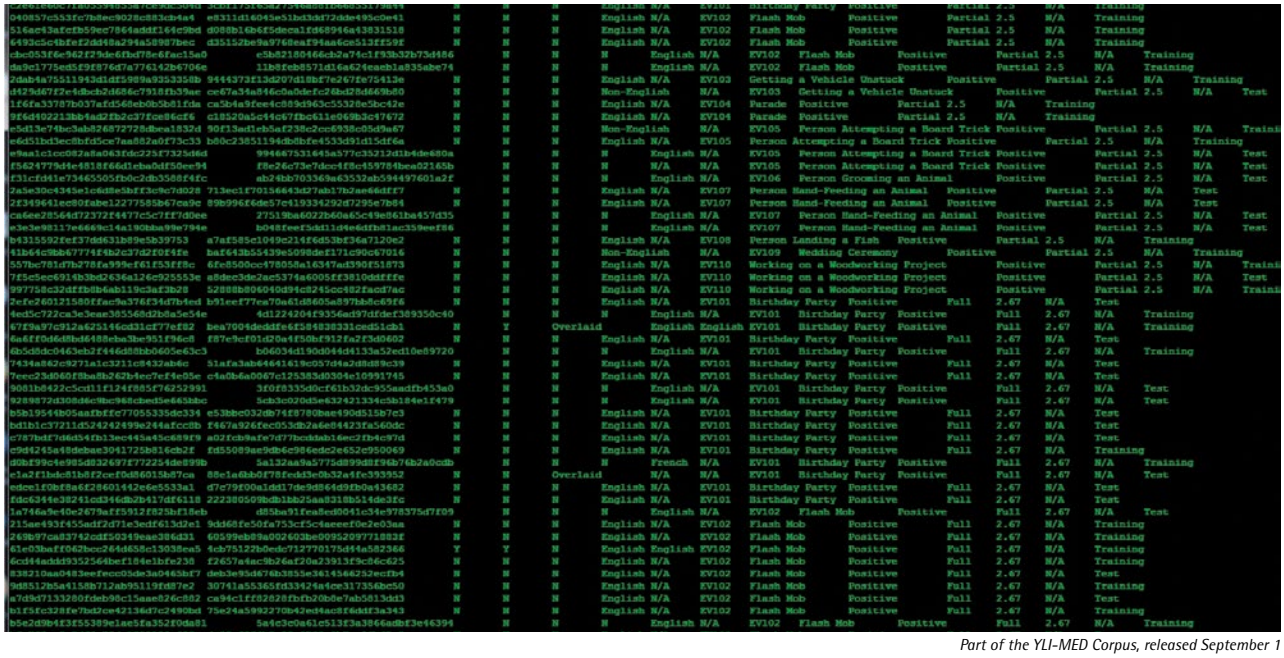
The Audio and Multimedia team is working with Yahoo Labs and Lawrence Livermore National Laboratory to process and analyze 100 million photos and videos publicly available under Creative Commons licenses. At more than 50 terabytes, the collection, the Yahoo Flickr Creative Commons 100 Million, is believed to be the largest corpus of user-generated multimedia publicly available for research.

On September 1, the team released full sets of three audio features for the nearly 800,00 videos in the corpus, computed on LLNL's Cray Catalyst supercomputer. Audio features, or statistical representations of sound, are used commonly in speech and audio recognition. The features released on September 1 are Mel frequency cepstral coefficients, Kaldi pitch, and subband autocorrelation classification. These can be used to build and train systems that automatically analyze the content of media with audio, such as videos.

The team also released an annotated index of more than 50,000 videos selected from the corpus. The index categorizes videos that depict 10 "events" based on events used in the 2011 multimedia Event Detection task of the National Institute of Standards and Technology's TRECVID Evaluation, which challenges participants to build systems that can find videos of the events. One of the events, for example, is "getting a vehicle unstuck." The new YLI-MED index released by the team lists 2,000 videos that depict events and 48,000 that do not, providing counter-examples needed to train systems.

Also on September 1, the team released a demo of audioCaffe, the first version of the framework being built for the SMASH project. The goal of the project is to create a media-analysis tool that can be run on parallel machines, reducing the time it takes to analyze media, and that can serve as a single framework for a variety of tasks like speech recognition, audio analysis, and video event detection. SMASH is funded by a National Science Foundation grant.

The set of features and the annotated indices are being held in the Yahoo-Lawrence-ICSI Corpus.



Part of the YLI-MED Corpus, released September 1

networking research at usenix security

Networking and Security researchers presented two papers at the USENIX Security Symposium, held in August in San Diego, examining two significant threats in modern cybersecurity: compromised browser extensions and the use of malware by oppressive governments.

William Marczak, a graduate student in the group, and Paxson worked with the Citizen Lab at the University of Toronto's Munk School of Global Affairs to investigate ways that the governments in Bahrain, Syria, and the United Arab Emirates use malware to identify and attack activists, journalists and others who have criticized their governments.

The researchers worked with activists to analyze the attacks, often initiated by social media and email messages masquerading as information about opposition movements. The researchers found that the attacks may have been a factor in setbacks to these movements ranging from public embarrassment to the criminal convictions of activists.

They looked at software marketed exclusively to governments, like Gamma International's FinSpy and Hacking Team's Remote Control System, that can record passwords, log keystrokes and take screenshots, among other capabilities. They also looked at the nongovernment-specific use of IP spy links, which can reveal the IP addresses of those who attempt to remain anonymous on social media, and remote access trojans.

Chris Grier, a senior researcher in Networking and Security, and Professor Vern Paxson, who leads the group, collaborated with researchers from UC Santa Barbara and UC San Diego to develop "Hulk," a program that identifies malicious code hidden in Google Chrome browser extensions.

Using data from the Chrome Web Store, they created an application that identifies security issues in popular extensions that expose users to malware and privacy invasion. Hulk is among the first of its kind and could lead to major security and policy changes in the web store as Google corrects the identified vulnerabilities.

icsi in the press

"'Shellshock' Bug Spells Trouble for Web Security," September 25, 2014, Brian Krebs, Krebs on Security

"New Security Flaw Could Be Worse than 'Heartbleed' Bug," September 25, 2014, Jana Katsuyama, KTVU News

"Worse than Heartbleed? Today's Bash Bug Could Break Security for Years," September 24, 2014, Russell Brandom, The Verge

"Shopping Online May Actually Be Safer Than Shopping In Person," September 15, 2014, Gerry Smith, The Huffington Post

"Devastating 'Heartbleed' Flaw Was Unknown Before Disclosure, Study Finds," September 10, 2014, Jeremy Kirk, PCWorld

"Heartbleed Attacks Started Within 24 Hours of Disclosure," September 10, 2014, Dan Worth, V3

"Research Finds No Large Scare Heartbleed Exploit Attempts Before Vulnerability," September 9, 2014, Dennis Fisher, Threatpost

"Dread Pirate Sunk by Leaky CAPTCHA," September 6, 2014, Brian Krebs, Krebs on Security

"Data: Nearly All U.S. Home Depot Stores Hit," September 3, 2014, Brian Krebs, Krebs on Security

"IPv6 Adoption Starting to Add Up to Real Numbers: 0.6 Percent," August 28, 2014, Iljitsch van Beijnum, Ars Technica

"IPv4 Is Not Enough," August 26, 2014, Marc Eisenbarth, Arbor Networks IT Security Blog

"What Do You Do With 100 Million Photos? David A. Shamma and the Flickr Photos Dataset," August 25, 2014, Trevor Owens, The Signal, Library of Congress

"Inside the Sneaky, Surprisingly Large World of Rogue Chrome Extensions," August 20, 2014, Jeremy Kirk, PCWorld

"Securobods Claim Middle East Govts' Fingerprints All Over Malware Flung at Journos," July 31, 2014, Darren Pauli, The Register

"Arab Monarchies Use Malware to Track Journalists," July 31, 2014, Joseph Marks, Politico

"Service Drains Competitors' Online Ad Budget," July 25, 2014, Brian Krebs, Krebs on Security

"Yahoo Releases Massive Flickr Dataset, and a Supercomputer Steps Up to Analyze It," July 3, 2014, Derrick Harris, GigaOM

"2014: The Year Extortion Went Mainstream," June 26, 2016, Brian Krebs, Krebs on Security

"How the FBI Brought Down Cyber-Underworld Site Silk Road," May 15, 2014, Donna Leinwand Leger, USA Today

"Backdoor in Call Monitoring, Surveillance Gear," May 14, 2014, Brian Krebs, Krebs on Security

"It's Crazy What Can Be Hacked Thanks to Heartbleed," April 28, 2014, Robert McMillan, Wired

"Study Finds No Evidence of Heartbleed Attacks Before the Bug Was Exposed," April 16, 2014, Nicole Perlroth, New York Times

"Why Heartbleed Is the Most Dangerous Security Flaw On The Web," April 8, 2014, Russell Brandom, The Verge

publications

R. Craven, R. Beverly, and M. Allman. Techniques for the Detection of Faulty Packet Header Modifications. Naval Postgraduate School Technical Report NPS-CS-14-002, March 2014

R. Craven, R. Beverly, and M. Allman. A Middlebox-Cooperative TCP for a Non End-to-End Internet. Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication (SIGCOMM 2013), Chicago, Illinois, August 2014

J. Czyz, M. Allman, J. Zhang, S. Iekel-Johnson, E. Osterweil, and M. Bailey. Measuring IPv6 Adoption. Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication (SIGCOMM 2013), Chicago, Illinois, August 2014

E. K. Dodge and M. R. L. Petruck. Representing Caused Motion in Embodied Construction Grammar. Proceedings of the 2014 ACL Workshop on Semantic Parsing, Baltimore, Maryland, pp. 39-44, June 2014

J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. Proceedings of the 31st International Conference in Machine Learning (ICML), Beijing, China, June 2014

J. Feldman and S. Narayanan. Affordances, Actionability, and Simulation. Presented at the First Workshop on Affordances: Affordances in Vision for Cognitive Robotics, held in conjunction with Robotics Science and Systems 2014 (RSS 2014), Berkeley, California, July 2014

J. Feng, S. Jegelka, S. Yan, and T. Darrell. Learning Scalable Discriminative Dictionary with Sample Relatedness. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, Ohio, June 2014

E. Friedman, A. Ghodsi, and C.-A. Psomas. Strategyproof Allocation of Discrete Jobs on Multiple Machines. Proceedings of the Fifteenth ACM Conference on Economics and Computation (EC'14), pp. 529-546, Palo Alto, California, June 2014

E. J. Friedman, K. Young, D. Asif, I. Jutla, M. Liang, S. Wilson, A. S. Landsberg, and N. Schuff. Directed Progression Brain Networks in Alzheimer's Disease: Properties and Classification. Brain Connectivity, Vol. 4, No. 5, pp. 384-393, June 2014

R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. Technical Report, Preprint: arXiv:1311.2524, June 2014

D. F. Gleich and M. W. Mahoney. Anti-Differentiating Approximation Algorithms: A Case Study with Min-Cuts, Spectral, and Flow. Proceedings of the 31st International Conference in Machine Learning (ICML), Beijing, China, June 2014

D. Goehring, J. Hoffman, E. Rodner, K. Saenko, and T. Darrell. Interactive Adaptation of Real-Time Object Detectors. Proceedings of the IEEE International Conference in Robotics and Automation (ICRA), Hong Kong, China, May 2014

S. Guadarrama, E. Rodner, K. Saenko, N. Zhang, R. Farrell, J. Donahue, and T. Darrell. Open-Vocabulary Object Retrieval. Proceedings of the 10th Annual Conference on Robotics: Science and Systems (RSS X), Berkeley, California, July 2014

K. F. Heppin and M. R. L. Petruck. Encoding of Compounds in Swedish FrameNet. Proceedings of The 10th Workshop on Multiword Expressions (MWE 2014) Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014), Gothenburg, Sweden, April 2014

J. Hoffman, T. Darrell, and K. Saenko. Continuous Manifold Based Adaptation for Evolving Visual Domains. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, Ohio, June 2014, June 2014

J. Hoffman, S. Guadarrama, E. Tzeng, J. Donahue, R. Girshick, T. Darrell, and K. Saenko. Large Scale Detector Adaptation. Technical Report, Preprint: arXiv:1407.5035, August 2014

J. van Hout, L. Ferrer, D. Vergyri, N. Scheffer, Y. Lei, V. Mitra, and S. Wegmann. Calibration and Multiple System Fusion for Spoken Term Detection Using Linear Logistic Regression. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014), Florence, Italy, May 2014

F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer. DenseNet: Implementing Efficient ConvNet Descriptor Pyramids. Technical Report, Preprint: arXiv:1404.1869, April 2014

L. G. S. Jeub, P. Balachandran, M. A. Porter, P. J. Mucha, and M. W. Mahoney. Think Locally, Act Locally: The Detection of Small, Medium-Sized, and Large Communities in Large Networks. Technical Report, Preprint: arXiv:1403.3795, March 2014

A. Kapravelos, C. Grier, N. Chachra, C. Kruegel, G. Vigna and V. Paxson. Hulk: Eliciting Malicious Behavior in Browser Extensions. Proceedings of the 23rd USENIX Security Symposium, San Diego, California, August 2014

S. Karayev, M. Fritz, and T. Darrell. Anytime Recognition of Objects and Scenes. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, Ohio, June 2014

publications, continued

P. Kay. Unary Phrase Structure Rules and the Cognitive Linguistics Lexical Linking Theory. Theoretical Linguistics, Vol. 40, Issue 1-2, Pages 149–163, ISSN (Online) 1613-4060, ISSN (Print) 0301-4428, DOI: 10.1515/tl-2014-0006, July 2014

H. Lei and N. Mirghafori. Broad Phonetic Classes for Speaker Verification with Noisy, Large-Scale Data. ICSI Technical Report TR-14-001, August 2014

H. Li, A. Ghodsi, M. Zaharia, S. Shenker, and I. Stoica. Reliable, Memory Speed Storage for Cluster Computing Frameworks. Technical Report UCB/EECS-2014-135, EECS Department, University of California, Berkeley, June 2014

A. Lukyanenko, I. Nikolaevskiy, D. Kuptsov, A. Gurtov, A. Ghodsi, and S. Shenker. STEM+: Allocating Bandwidth Fairly To Tasks. ICSI Technical Report TR-14-001, April 2014

W. R. Marczak, J. Scott-Railton, M. Marquis-Boire, and V. Paxson. When Governments Hack Opponents: A Look at Actors and Technology. Proceedings of the 23rd USENIX Security Symposium, San Diego, California, August 2014

G. Raskutti and M. W. Mahoney. A Statistical Perspective on Randomized Sketching for Ordinary Least-Squares. Technical Report, Preprint: arXiv:1406.5986, June 2014

R. Rasti, M. Murthy, and V. Paxson. Temporal Lensing and its Application in Pulsing Denial of Service Attacks. Technical Report No. UCB/EECS-2014-129, UC Berkeley, Berkeley, California, May 2014

M. Sargent, E. Blanton, and M. Allman. Modern Application Layer Transmission Patterns from a Transport Perspective. Proceedings of the 15th Passive and Active Measurement Conference (PAM 2011), Los Angeles, California, March 2014

M. Sargent and M. Allman. Performance Within A Fiber-To-The-Home Network. ACM SIGCOMM Computer Communications Review, Vol. 44, No. 3, pp. 23-30, July 2014

K. Schomp, T. Callahan, M. Rabinovich, and M. Allman. Assessing DNS Vulnerability to Record Injection. Proceedings of the 15th Passive and Active Measurement Conference (PAM 2011), Los Angeles, California, March 2014

C. Scott, A. Wundsam, B. Raghavan, Z. Liu, S. Whitlock, A. El-Hassany, A. Or, J. Lai, E. Huang, H. B. Acharya, K. Zarifis, and S. Shenker. Troubleshooting Blackbox SDN Control Software with Minimal Causal Sequences. Proceedings of the annual conference of the ACM Special Interest Group on Data Communication (SIGCOMM '14), pp. 395-406, Chicago, Illinois, August 2014

H. O. Song, R. Girshick, S. Jegelka, J. Mairal, Z. Harchaoui, and T. Darrell. On Learning to Localize Objects With Minimal Supervision. Technical Report, Preprint: arXiv:1403.1024, May 2014

H. O. Song, Y. J. Lee, S. Jegelka, and T. Darrell. Weakly-Supervised Discovery of Visual Pattern Configurations. Technical Report, Preprint: arXiv:1406.6507, June 2014

O. Vinyals and S. Wegmann. Chasing the Metric: Smoothing Learning Algorithms for Keyword Detection. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014), Florence, Italy, May 2014

N. Weaver, C. Kreibich, M. Dam, and V. Paxson. Here Be Web Proxies. Proceedings of the 15th International Conference on Passive and Active Measurement (PAM), Los Angeles, California, in Lecture Notes in Computer Science, Vol. 8362, pp. 183-192, March 2014

Y. Xiong, D. Scharstein, A. Chakrabarti, T. Darrell, B. Sun, K. Saenko, and T. Zickler. Modeling Radiometric Uncertainty for Vision with Tone-Mapped Color Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, Preprint. IEEE computer Society Digital Library. IEEE Computer Society, April 2014

J. Yang, V. Sindhwani, Q. Fan, H. Avron, and M. W. Mahoney. Random Laplace Feature Maps for Semigroup Kernels on Histograms. Proceedings of the IEEE Conference of Computer Vision and Pattern Recognition (CVPR), Columbus, Ohio, June 2014

J. Yang, V. Sindhwani, H. Avron, and M. W. Mahoney. Quasi-Monte Carlo Feature Maps for Shift-Invariant Kernels. Proceedings of the 31st International Conference in Machine Learning (ICML), Beijing, China, June 2014

N. Zhang, M. Paluri, M. Ranzato, T. Darrell, and L. Bourdev. PANDA: Pose Aligned Networks for Deep Attribute Modeling. Technical Report, Preprint: arXiv:1311.5591, May 2014

Owen Edwards, son of researcher Erik Edwards, and Michele Stecca, a research initiatives visitors, at ICSI's 2014 trip to see the Oakland Athletics baseball team. At one point in this year's game, a foul ball flew into ICSI's section. Stecca, who is from Italy and had never attended an American baseball game, grabbed the ball and handed it over to Owen, an avid baseball fan.