Roxana Geambasu

Computer Science Department, Columbia University
1214 Amsterdam Avenue, New York, NY 10027

roxana@cs.columbia.edu
http://www.cs.columbia.edu/~roxana/
US Permanent Resident (Green Card)

Education

Ph.D. in Computer Science, University of Washington (UW), August 2011.
Thesis title: "Empowering Users with Control over Cloud and Mobile Data."

M.S. in Computer Science, University of Washington, 2007.

B.S. in Computer Science, Polytechnic University of Bucharest, Romania (PUB), 2005.

Work Experience

Assistant Professor, Columbia University (CU), 2011 – present.

Software engineering intern, Google, summer 2008.

Research intern, Microsoft Research, summer 2007.

Awards

Early Career Award in Cybersecurity from the University of Washington Center of Academic Excellence, 2015.

Microsoft Faculty Fellowship, 2014.

Popular Science "Brilliant 10," 2014.

NSF CAREER Award, 2014.

Google Research Award, 2013.

Honorable mention for the inaugural SIGOPS Dennis M. Ritchie Dissertation Award, 2013.

The William Chan Memorial Dissertation Award, 2011.

Best Student Paper award at the European Conference on Computer Systems, 2011.

Outstanding Student Paper award at the 18^{th} USENIX Security Symposium, 2009.

Google Ph.D. Fellowship in Cloud Computing, 2009 – 2011.

Valedictorian of the 2005 Class at the Polytechnic University of Bucharest (PUB), 2005.

Research Theme

My research revolves around broad computer systems topics, including operating systems, distributed systems, and security and privacy. One overarching theme is the development of new abstractions for *rigorous and responsible data management*. I believe that many of the security and privacy challenges we face today stem from careless data management practices by mobile apps, service providers, and users. My collaborators and I are building a set of new operating and distributed systems abstractions that facilitate rigorous management of sensitive data.

Current Projects

Web transparency tools. Today's Web services leverage users' information – such as emails, search logs, or locations – and use them to target advertisements, prices, or products at users. Presently, users have little insight into how their data is used for such purposes. To enhance transparency, we are building a new set of tools system that detect what data – such as emails or searches – is used to target which ads in Gmail, which prices in Amazon, etc. The insight is to compare ads/prices witnessed by different accounts with similar, but not identical, subsets of the data [1,4].

Fairness testing tools for data-driven applications. Today's programmers routinely pass immense and varied kinds of personal data through increasingly complex machine learning algorithms, whose associations and inferences are difficult to anticipate and analyze. This results in a great risk for unwarranted associations, such as unintended discriminatory effects and racist labeling. We are building *FairTest*, a testing toolkit for programmers to discover these kinds of unintended consequences.

Modern protection abstractions for modern OSes. Data storage abstractions in OSes have evolved enormously. While traditional OSes used to provide fairly low-level abstractions – files and directories – modern OSes, including Android, iOS, OSX, and recent Windows, embed much higher-level abstractions, such as relational databases or object-relational models. Despite the change in abstraction, many crucial protection systems, such as encryption or deniable systems, still operate at the old file level, which often renders them ineffective. We are investigating new data protection abstractions that are more suitable for modern operating systems, including a new *logical data object* abstraction, which corresponds directly to user-level objects, such as emails, documents, or pictures [3,6].

Heterogeneous-database replication. We are building *Synapse*, a heterogeneous-database replication system, which lets programmers of complex, multi-service Web applications to share data across services running on very distinct database engines, in real time, and with solid consistency semantics. The key idea is to replicate data at model level by plugging into object-relational mappers, often used by modern Web applications. Synapse currently supports replication across ten different databases – more than any other heterogeneous-database replication system. It has been running at a NYC-based startup for more than a year [2].

Relevant Recent Projects

Virtual machine migration. We are improving virtual machine (VM) migration mechanisms by incorporating past state access histories [5] and hints provided by the guest operating system (in submission). For example, we show that state access histories can enable streaming of large VMs over wide area and even cellular networks with little loss in the VM's interactivity [5].

Keypad: Auditing file system for mobile devices. With today's limited anti-theft tools, users can neither assuredly restrict nor remotely monitor a thief's data accesses on a stolen or lost mobile device. I built Keypad, a new file system that enhances data security on mobile devices by providing users with post-theft fine-grained access auditing [9].

Comet: Cloud storage customization with active storage. Today's cloud storage services, such as Amazon S3, are highly inflexible and impose a variety of constraints on their clients: specific data consistency properties, fixed replication factors, limited logging, etc. I built Comet, an extensible storage service that allows clients to inject snippets of code that control the behavior of their data inside the storage service [11].

Vanish: Data lifetime control with self-destructing data. Users' migration to cloud and Web services is causing them to lose control over the lifetime of their data. Vanish is a self-destructing data system that allows users to impose timeouts on their Web data, such as emails, Facebook messages, or Google Docs [10,12], `http://vanish.cs.washington.edu`.

Publications

[1] Mathias Lecuyer, Riley B. Spahn, Giannis Spiliopoulos, Augustin Chaintreau, Roxana Geambasu, and Daniel Hsu. "Sunlight: Fine-grained Targeting Detection at Scale with Statistical Confidence." In *Proceedings of the ACM Conference on Computer and Communications Security (CCS)*, Denver, Colorado, October 2015.

[2] Nicolas Viennot, Mathias Lecuyer, Jonathan Bell, Roxana Geambasu, and Jason Nieh. "Synapse: New Data Integration Abstractions for Agile Web Application Development." In *Proceedings of Proceedings of the European Conference on Computer Systems (EuroSys)*, Bordeaux, France, April 2015.

[3] Riley Spahn, Jonathan Bell, Sravan Bhamidipati, Michael Lee, Roxana Geambasu, and Gail Kaiser. "Pebbles: Fine-Grained Data Management Abstractions for Modern Operating Systems." In *Proceedings of USENIX Operating Systems Design and Implementation (OSDI)*, Broomfield, CO, October 2014.

[4] Mathias Lécuyer, Guillaume Ducoffe, Francis Lan, Andrei Papancea, Theofilos Petsios, Riley Spahn, Augustin Chaintreau, and Roxana Geambasu. "XRay: Increasing the Web's Transparency with Differential Correlation." In *Proceedings of the USENIX Security Symposium*, San Diego, CA, August 2014.

[5] Yoshihisa Abe, Roxana Geambasu, Kaustubh Joshi, H. Andres Lagar-Cavilla, and Mahadev Satyanarayanan. "vTube: Efficient Streaming of Virtual Appliances Over Last-Mile Networks." In *Proceedings of the ACM Symposium on Cloud Computing (SoCC)*, Santa Clara, CA, October 2013.

[6] Yang Tang, Phillip Ames, Sravan Bhamidipati, Ashish Bijlani, Roxana Geambasu, and Nikhil Sarda. "CleanOS: Limiting mobile data exposure with idle eviction." In *Proceedings of the 10th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, Hollywood, CA, October 2012.

[7] Angelos D. Keromytis, Roxana Geambasu, Simha Sethumadhavan, Salvatore J. Stolfo, Junfeng Yang, Azzedine Benameur, Marc Dacier, Matthew Elder, Darrell Kienzle, and Angelos Stavrou. "The MEERKATS Cloud Security Architecture." In *Proceedings of the 3rd International Workshop on Security and Privacy in Cloud Computing (ICDCS-SPCC)*, Macao, China, June 2012.

[8] Roxana Geambasu. "Regaining Control over Cloud and Mobile Data." Ph.D. dissertation, University of Washington, Seattle, WA, August 2011. **Honorable Mention for the Inaugural Dennis M. Ritchie Doctoral Dissertation Award (2013). The William Chan Memorial Dissertation Award (2011).**

[9] Roxana Geambasu, John P. John, Tadayoshi Kohno, Steven D. Gribble, and Henry M. Levy. "Keypad: An auditing file system for theft-prone devices." In *Proceedings of the European Conference on Computer Systems (EuroSys)*, Salzburg, Austria, April 2011. **Best Student Paper Award.**

[10] Roxana Geambasu, Tadayoshi Kohno, Arvind Krishnamurthy, Amit Levy, Henry M. Levy, Paul Gardner, and Vinnie Moscaritolo. "New directions for self-destructing data." Technical Report, University of Washington, UW-CSE-11-08-01, 2011.

[11] Roxana Geambasu, Amit Levy, Tadayoshi Kohno, Arvind Krishnamurthy, and Henry M. Levy. "Comet: An active distributed key/value store." In *Proceedings of the* 9^{th} *USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, Vancouver, Canada, October 2010.

[12] Roxana Geambasu, Tadayoshi Kohno, Amit Levy, and Henry M. Levy. "Increasing data privacy with self-destructing data." In *Proceedings of the* 18^{th} *USENIX Security Symposium*, Montreal, Canada, August 2009. **Outstanding Student Paper Award.**

[13] Roxana Geambasu, Steven D. Gribble, and Henry M. Levy. "CloudViews: Communal data sharing in public clouds." In *Proceedings of the* 1^{st} *USENIX Workshop on Hot Topics in Cloud Computing (HotCloud)*, San Diego, CA, June 2009.

[14] Roxana Geambasu, Andrew Birrell, and John MacCormick. "Using formal specification to understand and compare fault-tolerant storage systems." In *Proceedings of the* 38^{th} *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN-DCCS)*, Anchorage, AK, June 2008.

[15] Roxana Geambasu, Cherie Cheung, Alexander Moshchuk, Steven D. Gribble, and Henry M. Levy. "Organizing and sharing of Web-service objects with Menagerie." In *Proceedings of the* 17^{th} *International World Wide Web Conference (WWW)*, Beijing, China, April 2008.

[16] Roxana Geambasu, Magdalena Balazinska, Steven D. Gribble, Henry M. Levy. "HomeViews: Peer-to-peer middleware for personal data sharing applications." In *Proceedings of the* 26^{th} *ACM International Conference on Management of Data (SIGMOD)*, Beijing, China, June 2007.

[17] Roxana Geambasu, Tanya Bragin, Jaeyeon Jung, Magdalena Balazinska. "On-Demand view materialization and indexing for network forensic analysis." In *Proceedings of the* 3^{rd} *International Workshop on Networking Meets Databases (NetDB)*, Boston, MA, April 2007.

Impact

Invited panelist for an event organized by the National Academy of Sciences on Privacy for the Intelligence Community, 2015. Attendees included members from the Intelligence Community and academic and industrial experts who discussed about emerging technologies challenge the Intelligence Community and its ability to manage user privacy.

Invited to talk about my work on data privacy and transparency at the Federal Trade Commission, 2015.

Invited to talk about my work on data privacy and transparency on Capitol Hill, 2014. Attendees include members of the Cybersecurity Caucus of the House of Representatives.

New York Times Bits article about the XRay project, 2014. `http://bits.blogs.nytimes.com/2014/08/18/xray-a-new-tool-for-tracking-the-use-of-personal-data-on-the-web/`. Other media coverage followed, including participation in the Lopate Show at the NYC branch of NPR.

Deployment in production of an early prototype of Synapse, a data sharing system, at Crowdtap, a NYC-based startup, 2013. The system has been running for about a year with ≈650M users. `https://github.com/crowdtap/promiscuous`.

First deployment of security measures in a commercial, giant-scale distributed hash table (DHT), 2010. I designed, evaluated, and deployed practical security defenses against Sybil data-crawling attacks on DHTs. My challenging experience with this deployment led to the design of Comet, an extensible DHT, published in OSDI 2010.

Extensive media coverage around the Vanish project, 2009. Articles appeared in The New York Times, NPR, The Economist, King 5 Seattle, and other outlets.

Professional Service

Co-organizer of ISAT Workshop "Whither the Data" on understanding complex flows in data ecosystems, 2015.

PC *USENIX Security Symposium*, 2015.

PC *European Conference on Computer Systems (EuroSys)*, 2015.

Member of Information Science and Technology (ISAT) study group, a select advisory group to DARPA that provides advice on long range research directions in information sciences and technology, 2014-2017.

NSF Panel for Secure and Trustworthy Computing (SaTC), 2014.

PC *USENIX Operating Systems Design and Implementation Conference (OSDI)*, 2014.

PC *ACM Cloud Computing Security Workshop (CCSW)*, 2013.

PC *ACM Symposium on Cloud Computing (SoCC)*, 2013.

PC *USENIX Security Symposium*, 2013.

PC *Workshop on Hot Topics in Operating Systems (HotOS)*, 2013.

PC *Workshop on Mobile Computing Systems and Applications (HotMobile)*, 2013.

PC *European Conference on Computer Systems (EuroSys)*, 2013.

Review Committee for the 2013 EuroSys Roger Needham Ph.D. Award, 2013.

NSF Panel for Computer Systems Research (CSR), 2012.

External PC *USENIX Operating Systems Design and Implementation (OSDI)*, 2012.

PC *USENIX Security Symposium*, 2012.

PC *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2011.

Diversity, Outreach

Organize hands-on app programming workshop for the Annual Engineering Exploration event organized by the Society of Women Engineers for New York City K-12 female students, 2014.

Faculty advisor on Columbia's Engleston Scholar mentorship program, 2013.

Department host at Columbia's Engineering Women's Forum, 2012, 2013.

Panelist for New York City Girls Computer Science and Engineering Conference, 2012.

Funding

PI on NSF SaTC Medium, "Scalable Web Transparency: New Scientific Building Blocks, Tools, and Measurements to Tame the Data-Driven Web," $1,588,998, 2015-2019.

Microsoft Faculty Fellowship, $200,000, 2014.

Microsoft Research gift, $15,000, 2014.

PI on NSF CAREER, "New Operating Systems Abstractions for Responsible Data Management," $499,999, 2014-2019.

PI on Google Faculty Fellowship, "Promiscuous: Scalable, Consistent Firehose for Data-Driven Web Service Integrations," $79,807 across two Co-PIs, 2013-2014.

PI on Columbia Provost's Grant for Junior Faculty Who Contribute to the Diversity Goals of the University, "CleanOS: Limiting Sensitive Data Exposure in Mobile Operating Systems," $25,000, 2013-2014.

PI on DARPA Contract No. FA8650-11-C-7190, "MEERKATS: Maintaining EnterprisE Resiliency via Kaleidoscopic Adaptation & Transformation of Software Services," $6,619,270 across six Co-PIs, 2011-2015. Original PI was Angelos Keromytis.

Invited Talks

"Increasing Privacy in a Data-Driven World." Panelist introductory talk at National Academies Workshop on Privacy for the Intelligence Community, 2015.

"Synapse: A Microservices Architecture for Heterogeneous-Database Web Applications." Invited Speaker at Workshop on Cloud Programmability co-located with the Microsoft Faculty Summit, 2015.

"Increasing Privacy in a Data-Driven World." Federal Trade Commission, 2015.

"Increasing Privacy in a Data-Driven World." Microsoft Research NYC, 2015.

"Increasing Privacy in a Data-Driven World." Columbia Womensphere Innovation Summit organized by the Womensphere Foundation and Columbia Graduate Society of Women Engineers, 2015.

"Increasing Privacy in a Data-Driven World." Columbia Undergraduate Scholars Program, 2015.

Panelist for the "Princeton Web Transparency Conference," 2014.

"Toward a Transparent Web." Invited speaker at the Diversity workshop co-located with OSDI, 2014.

"New Abstractions for Responsible Big-Data Management." Invited speaker at the DIMACS Workshop on Secure Cloud Computing, 2014.

"Increasing Privacy in a Data-Driven World." Microsoft Faculty Fellowship final competition, 2014.

"New Abstractions for Responsible Big-Data Management." Microsoft Research, 2013.

"Responsible Big-Data Management." Journalism Security Seminar organized by Columbia's Journalism School, 2013.

"Regaining Control over Mobile and Cloud Data." Cloud Computing Security Forum, part of the IEEE Global Communications Conference (Globecom), 2011.

"Cloud Computing: Benefits and Challenges." Expert talk at Columbia Senate Information Technology Committee, 2011.

"Regaining Control over Mobile and Cloud Data." Invited talk delivered at universities and industrial labs: AT&T Labs NYC, Brown University, Carnegie Mellon University, Columbia University, Cornell University, Duke University, Georgia Institute of Technology, Google, Harvard University, IBM Research, Intel Corporation, Massachussetts Institute of Technology (MIT), Microsoft Research, New York University, Symantec Research Labs, University of California at Los Angeles, University of Southern California, 2011-2012.

"Self-destructing Data and Beyond." Invited talk at the University of British Columbia Systems Colloquium, 2010.

"Vanish: Increasing Data Privacy with Self-destructing Data." Invited talk at the Google Graduate Student Forum, 2010.