# Roxana Geambasu

Department of Computer Science
Columbia University
500 W. 120th St., Mailcode: 0401
New York, NY 10027

Phone: +1 (212) 939-7099
E-mail: roxana@cs.columbia.edu
WWW: http://www.cs.columbia.edu/~roxana/

## Research Interests

Software systems with a focus on security and privacy problems.

## Education

**Ph.D. Computer Science, University of Washington**, August 2011.

Advisors: Prof. Henry M. Levy, Tadayoshi Kohno, Steven D. Gribble.
Dissertation title: Empowering Users with Control over Cloud and Mobile Data.

**M.S. Computer Science, University of Washington**, June 2007.

**B.S. Computer Science and Engineering, Polytechnic University of Bucharest, Romania**, May 2005.

Valedictorian of the 2005 Computer Science and Engineering class.

## Positions

**Department of Computer Science, Columbia University**, New York, NY.

*Associate professor with tenure*, January 2020 to present.

*Associate professor*, January 2016 to January 2020.

*Assistant professor*, July 2012 to December 2015.

*Instructor*, July 2011 to June 2012.

**Meta, Inc.**, New York, NY.

*Contingent worker*, January 2024.

**Google, Inc.**, New York, NY.

*Visiting researcher*, February 2021 to December 2021.

**Department of Computer Science and Engineering, University of Washington**, Seattle, WA.

*Graduate student researcher*, September 2005 to August 2011.

**Google, Inc.**, Seattle, WA.

*Software engineering intern*, June 2008 - August 2008.

**Microsoft Research**, Mountain View, CA.

*Research intern*, June 2007 - August 2007.

**Department of Computer Science and Engineering, Polytechnic University of Bucharest**, Romania.

*Undergraduate student researcher*, September 2003 to June 2005.

# Awards and Honors

Google Research Award, 2019.

Alfred P. Sloan Faculty Fellowship, 2016.

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

Microsoft Faculty Fellowship, 2014.

Popular Science Brilliant 10, 2014.

NSF CAREER Award, 2014.

Elected member of DARPA's Information Science and Technology (ISAT) study group, 2014-2017.

Google Research Award, 2013.

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

The William Chan Memorial Dissertation Award, 2011.

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

Best Paper award at the USENIX Security Symposium, 2009.

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

Valedictorian of the 2005 Computer Science Class at Polytechnic University of Bucharest, 2005.

# Publications

Underline indicates students for whom I am the primary advisor.

## *Refereed Conference Proceedings*

1. Kelly Kostopoulo, Pierre Tholoniat, Asaf Cidon, Roxana Geambasu, and Mathias Lecuyer. Turbo: Effective caching in differentially-private databases. In *Proceedings of the ACM Symposium on Operating Systems Principles (SOSP)*, 2023.

2. Matthew Jagielski, Stanley Wu, Alina Oprea, Jonathan Ullman, and Roxana Geambasu. How to combine membership-inference attacks on multiple updated models. In *Proceedings of the Privacy-Enhancing Technologies Symposium (PETS)*, 2023.

3. Tao Luo, Mingen Pan, Pierre Tholoniat, Asaf Cidon, Roxana Geambasu, and Mathias Lecuyer. Privacy budget scheduling and orchestration. In *Proceedings of the USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2021.

4. Mathias Lecuyer, Riley Spahn, Kiran Vodrahalli, Roxana Geambasu, and Daniel Hsu. Privacy accounting and quality control in the Sage differentially private machine learning platform. In *Proceedings of the ACM Symposium on Operating Systems Principles (SOSP)*, 2019.

5. Mathias Lecuyer, Vaggelis Atlidakis, Roxana Geambasu, Daniel Hsu, and Suman Jana. Certified robustness to adversarial examples with differential privacy. In *Proceedings of the IEEE Security and Privacy Symposium (IEEE S&P)*, 2019.

6. Mathias Lecuyer, Riley Spahn, Roxana Geambasu, Tzu-Kuo Huang, and Siddhartha Sen. Pyramid: Enhancing selectivity in big data protection with count featurization. In *Proceedings of the IEEE Security and Privacy Symposium (IEEE S&P)*, 2017.

7. Florian Tramer, Vaggelis Atlidakis, Roxana Geambasu, Daniel Hsu, Jean-Pierre Hubaux, Mathias Humbert, Ari Juels, and Huang Lin. Fairtest: Discovering unwarranted associations in data-driven applications. In *Proceedings of the IEEE European Symposium on Security and Privacy (EuroS&P)*, 2017.

8. Yoshihisa Abe, Roxana Geambasu, Kaustubh Joshi, and Mahadev Satyanarayanan. Urgent virtual machine migration with enlightened post-copy. In *Proceedings of the Conference of Virtual Execution Environments (VEE)*, 2016.

9. Vaggelis Atlidakis, Jeremy Andrus, Roxana Geambasu, Dimitris Mitropoulos, and Jason Nieh. POSIX abstractions in modern operating systems: The old, the new, and the missing. In *Proceedings of the IEEE European Conference on Computer Systems (EuroSys)*, 2016.

10. Mathias Lecuyer, Riley Spahn, Yannis 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)*, 2015.

11. Nicolas Viennot, Mathias Lecuyer, Jonathan Bell, Roxana Geambasu, and Jason Nieh. Synapse: New data integration abstractions for agile web application development. In *Proceedings of the ACM European Conference on Computer Systems (EuroSys)*, 2015.

12. Riley Spahn, Jonathan Bell, Michael Lee, Sravan Bhamidipati, Roxana Geambasu, and Gail Kaiser. Pebbles: Fine-grained data management abstractions for modern operating systems. In *Proceedings of the USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2014.

13. Mathias Lecuyer, 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 USENIX Security*, 2014.

14. 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)*, 2013.

15. Yang Tang, Phillip Ames, Sravan Bhamidipati, Ashish Bijlani, Roxana Geambasu, and Nikhil Sarda. CleanOS: Mobile OS abstractions for managing sensitive data. In *Proceedings of the USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2012.

16. Roxana Geambasu, John P. John, Steven D. Gribble, Tadayoshi Kohno, and Henry M. Levy. Keypad: An auditing file system for theft-prone devices. In *Proceedings of the ACM European Conference on Computer Systems (EuroSys)*, 2011 **Award Paper**.

17. Roxana Geambasu, Amit Levy, Tadayoshi Kohno, Arvind Krishnamurthy, and Henry M. Levy. Comet: An active distributed key/value store. In *Proceedings of the USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2010.

18. Roxana Geambasu, Tadayoshi Kohno, Amit Levy, and Henry M. Levy. Vanish: Increasing data privacy with self-destructing data. In *Proceedings of USENIX Security*, 2009. **Award Paper**.

19. Roxana Geambasu, Cherie Cheung, Alexander Moshchuk, Steven D. Gribble, and Henry M. Levy. The organization and sharing of web-service objects with menagerie. In *Proceedings of the International World Wide Web Conference (WWW)*, 2008.

20. Roxana Geambasu, Magdalena Balazinska, Steven D. Gribble, and Henry M. Levy. Homeviews: Peer-to-peer middleware for personal data sharing applications. In *Proceedings of the CM International Conference on Management of Data (SIGMOD)*, 2007.

## *Refereed Workshop/Journal/Magazine Papers*

21. Rachel Cummings, Damien Desfontaines, David Evans, Roxana Geambasu, Yangsibo Huang, Matthew Jagielski, Peter Kairouz, Gautam Kamath, Sewoong Oh, Olga Ohrimenko, Nicolas Papernot, Ryan Rogers, Milan Shen, Shuang Song, Weijie Su, Andreas Terzis, Abhradeep Thakurta, Sergei Vassilvitskii, Yu-Xiang Wang, Li Xiong, Da Yu, Sergey Yekhanin, Huanyu Zhang, and Wanrong Zhang. Advancing differential privacy: where we are now and future directions for real-world deployment. *Harvard Data Science Review*, 6(1), 2024.

22. Mathias Lecuyer, Riley Spahn, Kiran Vodrahalli, Roxana Geambasu, and Daniel Hsu. Privacy accounting and quality control in the Sage differentially private ml platform. *ACM SIGOPS Operating Systems Review (OSR)*, 2019.

23. Mathias Lecuyer, Riley Spahn, Roxana Geambasu, Tzu-Kuo Huang, and Siddhartha Sen. Enhancing selectivity in big data. In *IEEE Security and Privacy Magazine*, 2018.

24. Vaggelis Atlidakis, Jeremy Andrus, Roxana Geambasu, Dimitris Mitropoulos, and Jason Nieh. POSIX has become outdated. In *USENIX ;login: Magazine*, 2016.

25. 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*, 2012.

26. Roxana Geambasu, Steven D. Gribble, and Henry M. Levy. Cloudviews: Communal data sharing in public clouds. In *Proceedings of the Workshop on Hot Topics in Cloud Computing (HotCloud)*, 2009.

27. Roxana Geambasu, John MacCormick, and Andrew Birrell. Experiences with formal specification of fault-tolerant file systems. In *Proceedings of the International Conference on Dependable Systems and Networks (DSN)*, 2008.

28. Roxana Geambasu, Tanya Bragin, Jaeyeon Jung, and Magdalena Balazinska. On-demand view materialization and indexing for network forensic analysis. In *Proceedings of the International Workshop on Networking Meets Databases (NetDB)*, 2007.

## *Technical Reports and Design Docs*

29. Benjamin Case, Ben Savage, Martin Thomson, Christian Berkhoff, Roxana Geambasu, Richa Jain, Alex Koshelev, Andy Leiserson, Daniel Masny, and Erik Taubeneck. Hybrid proposal design for patcg. Technical report, 2024.

30. Benjamin Case and Roxana Geambasu. Dp budgeting for hybrid proposal. Technical report, 2024.

31. Alexey Kurakin, Shuang Song, Steve Chien, Roxana Geambasu, Andreas Terzis, and Abhradeep Thakurta. Toward training at imagenet scale with differential privacy. Technical Report 2201.12328, arXiv, 2022.

32. Mathias Lecuyer, Riley Spahn, Kiran Vodrahalli, Roxana Geambasu, and Daniel Hsu. Privacy accounting and quality control in the Sage differentially private machine learning platform. Technical Report 1909.01502, arXiv, 2019.

33. Mathias Lecuyer, Vaggelis Atlidakis, Roxana Geambasu, Daniel Hsu, and Suman Jana. On the connection between differential privacy and adversarial robustness in machine learning. Technical Report 1802.03471v1, arXiv, 2018.

34. Mathias Lecuyer, Riley Spahn, Roxana Geambasu, Tzu-Kuo Huang, and Siddhartha Sen. Pyramid: Enhancing selectivity in big data protection with count featurization. Technical Report 1705.07512, arXiv, 2017.

35. Florian Tramer, Vaggelis Atlidakis, Roxana Geambasu, Daniel Hsu, Jean-Pierre Hubaux, Mathias Humbert, Ari Juels, and Huang Lin. Discovering unwarranted associations in data-driven applications with the fairtest testing toolkit. Technical Report 1510.02377v2, arXiv, 2017.

36. Roxana Geambasu, Tadayoshi Kohno, Arvind Krishnamurthy, Amit Levy, Henry M. Levy, Paul Gardner, and Vinnie Moscaritolo. New directions for self-destructing data systems. Technical Report UW-CSE-11-08-01, University of Washington, 2010.

### *Dissertations*

37. Roxana Geambasu. *Regaining Control over Cloud and Mobile Data*. PhD thesis, University of Washington, 2011.

### *Software*

Open-source code release of *PrivateKube*, Kubernetes-based privacy budget scheduling, 2019. `https://github.com/columbia/PrivateKube`.

Open-source code release of *PixelDP*, certified defense against adversarial examples, 2019. `https://columbia.github.io/pixeldp/`.

Open-source code release of *Pyramid*, selective data protection system, 2017. `https://columbia.github.io/pyramid/`.

Open-source code release of *FairTest*, fairness testing tool, 2017. `https://columbia.github.io/fairtest/`.

Open-source code release of *Sunlight*, our second-generation web transparency system, 2015. `https://columbia.github.io/sunlight/`.

Open-source code release of *XRay*, our first web transparency system, 2014. `http://xray.cs.columbia.edu`.

Open-source code release of *Pebbles*, an object-level protection system for persistent data, 2014. `https://github.com/columbia/pebbles_platform_dalvik`.

Open-source code release of *Synapse*, a heterogeneous-database integration system, 2013. `https://github.com/nviennot/synapse`. The code was adopted by Crowdtap, a New York City-based social media startup in 2013 and has run for at least two years in production with 650K users.

Open-source code release of *Vanish*, self-destructing data system, 2009. `https://vanish.cs.washington.edu/`. Security measures against Sybil attacks that we developed as part of this project were adopted by the Vuze commercial peer-to-peer file sharing system in 2010.

## Funding

Senior personnel & research thrust lead on the *NSF* ERC Center for Smart Streetscapes (CS3) (2022-2027, $25M). Main PI: Prof. Andrew Smyth, Columbia Civil Engineering. Geambasu's role: security and privacy research thrust lead and participant. Geambasu's share: 1 student for 5 years.

Co-PI for *DoE* award (2020-2021, $800,000). "Exascale Privacy-Preserving AI." Primary institution is Brookhaven National Lab, Columbia is a subcontractor. Geambasu's share: $101,197.

PI for *Google Research* award (2019-2020, $88,959). "Certified Robustness to Adversarial Examples with Differential Privacy." Sole PI.

PI for *Schmidt Futures* award (2018-2019, $100,000). "Enabling Small-Data Medical Research with Private Transferrable Knowledge from Big Data." Co-PIs: Nicholas Tatonetti and Daniel Hsu. Geambasu's share: $50,000.

PI for *Columbia Data Sciences Institute Seed Grant* (2018-2020, $200,000). "Enabling Small-Data Medical Research with Private Transferrable Knowledge from Big Data." Co-PIs: Nicholas Tatonetti and Daniel Hsu. Geambasu's share: $100,000.

PI for *Alfred P. Sloan Fellowship* (2016-2018, $55,000). "Privacy in a Data-Driven World." Sole PI.

PI for *NSF Secure and Trustworthy Computing* award (2015-2019, $1,588,998). "Scalable Web Transparency: New Scientific Building Blocks, Tools, and Measurements to Tame the Data-Driven Web." Co-PI: Augustin Chaintreau. Geambasu's share: $988,998.

PI for *Microsoft Faculty Fellowship* ($200,000, 2014-no end). Sole PI.

PI for *Microsoft Research* gift ($15,000, 2014-no end). Sole PI.

PI for *NSF CAREER* award ($499,999, 2014-2019). "New Operating Systems Abstractions for Responsible Data Management." Sole PI.

PI for *Google Research* award ($79,807, 2013-2014). "Promiscuous: Scalable, Consistent Firehose for Data-Driven Web Service Integrations." Co-PI: Jason Nieh. Geambasu's share: $39,903.

PI for *Columbia Provost's Grant for Junior Faculty* ($25,000, 2013-2014). "CleanOS: Limiting Sensitive Data Exposure in Mobile Operating Systems." Sole PI.

PI for *DARPA Mission-Resilient Clouds* contract ($6,619,270, 2011-2015). "MEERKATS: Maintaining EnterprisE Resiliency via Kaleidoscopic Adaptation & Transformation of Software Services." Co-PIs: Angelos Keromytis (was original PI), Salvatore Stolfo, Simha Sethumadhavan, Junfeng Yang, Matthew Elder, and Angelos Stavrou. Geambasu's share: $900,000.

## Invited Talks

World Wide Web Consortium (W3C) Private Advertising Technology Working Group (PATCG), invited speaker, "Differentially Private Budgeting for Hybrid IPA-PAM Proposal," 2024.

W3C Private Advertising Technology Working Group (PATCG), invited speaker, "Efficient On-device Differentially Private Budgeting," 2024.

Berkeley Sky Seminar Series, invited speaker, "Managing Privacy in Data-driven Workloads," 2023.

Boston University - RedHat Collaboratory Colloquium, invited speaker, "Managing Privacy in Data-driven Workloads," 2023.

Google PARADE (Privacy in Advertising) Workshop, invited speaker, "Differential Privacy in ML Infrastructure," 2022.

Boston-area Differential Privacy Seminar, "Security and Privacy Guarantees in Machine Learning with Differential Privacy," 2021.

Max Planck Institute for Software Systems, Distinguished Lecture Series, "Security and Privacy Guarantees in Machine Learning with Differential Privacy," 2020.

University of Maryland Distinguished Lectures Series, "Security and Privacy Guarantees in Machine Learning with Differential Privacy," 2019.

University of Washington security seminar, "Security and Privacy Guarantees in Machine Learning with Differential Privacy," 2019.

Microsoft Research security seminar, "Security and Privacy Guarantees in Machine Learning with Differential Privacy," 2019.

Princeton systems seminar, "Security and Privacy Guarantees in Machine Learning with Differential Privacy," 2019.

MIT security seminar, "Security and Privacy Guarantees in Machine Learning with Differential Privacy," 2019.

Northeastern University systems seminar, "Security and Privacy Guarantees in Machine Learning with Differential Privacy," 2019.

CMU systems + cybersecurity seminar, "Security and Privacy Guarantees in Machine Learning with Differential Privacy," 2019.

Army Strategic Planning Workshop organized by CMU on Autonomy and AI in Cybersecurity, "AI with Guarantees from Differential Privacy," 2019.

Berkeley RISELab seminar, "Security and Privacy Guarantees in Machine Learning with Differential Privacy," 2019.

Google Faculty Summit, "Differential Privacy Foundations for Security and Privacy in ML," 2019.

Harvard systems + programming languages seminar, "Security and Privacy Guarantees in Machine Learning with Differential Privacy," 2019.

University of Pennsylvania systems seminar, "Security and Privacy Guarantees in Machine Learning with Differential Privacy," 2019.

Stanford computer security seminar, "Security and Privacy Guarantees in Machine Learning with Differential Privacy," 2019.

Columbia Data Innovation Network Summit, "Differential Privacy as A Bedrock for Secure and Private ML," 2019.

TwoSigma, "Certifying Robustness to Adversarial Examples with Differential Privacy Theory," 2019.

New York City Women in Data Science (WiDS), "Privacy in a Data-Driven World," 2018.

Google Faculty Summit, "Selective Data Systems," 2017.

Hearst Corporation, "Privacy in a Data-Driven World," 2016.

Harvard University systems seminar, "Privacy in a Data-Driven World," 2016.

MIT CSAIL systems seminar, "Privacy in a Data-Driven World," 2016.

New York City Media Lab, "Privacy in a Data-Driven World," 2016.

Innovation and the Value of Privacy Conference, Columbia Business School, "Privacy in a Data-Driven World," 2016.

Federal Trade Comission (FTC) PrivacyCon conference, "Web Transparency at Scale," 2016.

Keynote at the Neural Information Processing Systems (NIPS) workshop on Learning Systems, "Privacy in a Data-Driven World," 2015.

TwoSigma, "Privacy in a Data-Driven World," 2015.

Stanford's computer security seminar, "Privacy in a Data-Driven World," 2015.

Berkeley's AMPLab seminar, "Privacy in a Data-Driven World," 2015.

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

Federal Trade Commission (FTC), "Transparency Infrastructures for the Data-Driven Web," 2015. Attendees included one of the FTC Commissioners and her staff.

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

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

NSF informational session on Capitol Hill, "Transparency in a Data-Driven World," 2014. Attendees included members of the Cybersecurity Caucus of the House of Representatives.

Princeton Web Transparency Conference, "Toward a Transparent Web," 2014.

Keynote at the Diversity Workshop co-located with OSDI, "Toward a Transparent Web," 2014.

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

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

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

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

Invited talk, titled "Regaining Control over Mobile and Cloud Data," delivered at multiple 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." University of British Columbia Systems Colloquium, 2010.

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

# Educational Activities

## *New Curriculum Development*

I developed Columbia University's distributed systems curriculum. It includes two courses:

**Distributed Systems Fundamentals (COMS 4113)**: Undergraduate-level course on distributed systems design and implementation, one of the few such courses in the country that is not centered around research paper reading. Topics include: distributed communication models, distributed synchronization, distributed file systems, replication, consistency models, failure models and fault tolerance, the consensus problem and consensus protocols (Paxos and RAFT), distributed transactions, large-scale batch infrastructures, streaming engines. In addition, the course reviews the design and implementation of several large-scale systems that are used in production at big companies. Lecture notes and materials are available online at: `https://columbia.github.io/ds1-class`.

**Advanced Distributed Systems (COMS 6114)**: Graduate-level research seminar on both foundational and cutting-edge, recent research in distributed systems. The seminar is centered around review and detailed discussion of research papers, along with a semester-long project for hands-on experience with some distributed systems research topic. Reading list is available online at: `https://columbia.github.io/ds2-class`.

**Private Systems (COMS 6995)**: Graduate-level research seminar that discusses the ethical and legal responsibilities of data scientists who are stewards of user data, along with a set of technologies that can be used to enhance privacy, accountability, fairness, and data protection in big data systems. A particular focus (and unique aspect) of this class will be to look at these technologies with a systems perspective of incorporating them into real data infrastructure systems. Lecture notes and materials are available online at: `https://columbia.github.io/private-systems-class/`.

## *Teaching Experience*

**Private Systems (6998)**, Columbia University, Spring 2024.

Enrollment: 16. Course evaluation: in progress. Instructor evaluation: in progress.

**Distributed Systems Fundamentals (COMS 4113)**, Columbia University, Fall 2023.

Enrollment: 106. Course evaluation: 4.26/5. Instructor evaluation: 4.26/5.

**Private Systems (6998)**, Columbia University, Spring 2023.

Enrollment: 20. Course evaluation: 5/5 (2 responders). Instructor evaluation: 5/5 (2 responders).

**Distributed Systems Fundamentals (COMS 4113)**, Columbia University, Fall 2022.

Enrollment: 108. Course evaluation: 4.35/5. Instructor evaluation: 4.42/5.

**Distributed Systems Fundamentals (COMS 4113)**, Columbia University, Spring 2022.

Enrollment: 186. Course evaluation: 4.23/5. Instructor evaluation: 4.14/5.

**Distributed Systems Fundamentals (COMS 4113)**, Columbia University, Fall 2020.

Enrollment: 198. Course evaluation: 4.05/5. Instructor evaluation: 4.18/5.

**Private Systems (6998)**, Columbia University, Spring 2020.

Enrollment: 3. Course evaluation: Not taken due to pandemic disruption. Instructor evaluation: Not taken due to pandemic disruption.

**Distributed Systems Fundamentals (COMS 4113)**, Columbia University, Fall 2019.

Enrollment: 143. Course evaluation: 4.2/5. Instructor evaluation: 4.18/5.

**Advanced Distributed Systems (COMS 6114)**, Columbia University, Spring 2019.

Enrollment: 8. Course evaluation: 5./5. Instructor evaluation: 5./5.

**Distributed Systems Fundamentals (COMS 4113)**, Columbia University, Fall 2018.

Enrollment: 118. Course evaluation: 3.7/5. Instructor evaluation: 3.5/5.

**Advanced Distributed Systems (COMS 6114)**, Columbia University, Spring 2018.

Enrollment: 10. Course evaluation: 4.0/5. Instructor evaluation: 4.0/5.

**Computer Systems for Big Data (W4212)**, Columbia University, Spring 2018. Co-instructors: Eugene Wu (Columbia) and Sambit Sahu (IBM).

Enrollment: 158. Course evaluation: 2.65/5. Instructor evaluation: 3.3/5.

**Distributed Systems Fundamentals (COMS 4113)**, Columbia University, Fall 2017.

Enrollment: 73. Course evaluation: 3.83/5. Instructor evaluation: 3.88/5.

**Computer Systems for Big Data (W4212)**, Columbia University, co-instructor, Spring 2016. Co-instructors: Eugene Wu (Columbia) and Sambit Sahu (IBM). Enrollment: 79. Course evaluation: 3.8/5. Instructor evaluation: 3.97/5.

**Advanced Distributed Systems (COMS 6114)**, Columbia University, Fall 2015.

Enrollment: 14. Course evaluation: 3.6/5. Instructor evaluation: 3.6/5.

**Advanced Distributed Systems (COMS 6114)**, Columbia University, Spring 2015.

Enrollment: 21. Course evaluation: 4.2/5. Instructor evaluation: 4.4/5.

**Computer Systems for Big Data (W4212)**, Columbia University, co-instructor, Spring 2015. Co-instructors: Simha Sethumadhavan (Columbia) and Li Erran Li (ATT). Enrollment: 18. Course evaluation: 3.9/5. Instructor evaluation: 3.9/5.

**Distributed Systems Fundamentals (COMS 4113)**, Columbia University, Fall 2014.
Enrollment: 41. Course evaluation: 3.7/5. Instructor evaluation: 3.7/5.

**Cloud and Mobile Challenges Seminar (COMS6998)**, Columbia University, Spring 2014.
Enrollment: 24. Course evaluation: 3.8/5. Instructor evaluation: 4.2/5.

**Distributed Systems Fundamentals (COMS 4113)**, Columbia University, Fall 2013.
Enrollment: 16. Course evaluation: 4.1/5. Instructor evaluation: 4.0/5.

**Cloud and Mobile Challenges Seminar (COMS6998)**, Columbia University, Spring 2013.
Enrollment: 24. Course evaluation: 3.9/5. Instructor evaluation: 4.3/5.

**Distributed Systems Fundamentals (COMS 4113)**, Columbia University, Fall 2012.
Enrollment: 16. Course evaluation: 4.2/5. Instructor evaluation: 4.2/5.

**Cloud and Mobile Challenges Seminar (COMS6998)**, Columbia University, Fall 2011.
Enrollment: 23. Course evaluation: not found. Instructor evaluation: not found.

## *Advising Experience*

**In-progress primary Ph.D. advising of students:**

Pierre Tholoniat: Columbia University; Ph.D. student 2019 to present.

**Completed Ph.D. dissertations sponsored as primary advisor:**

Vaggelis Atlidakis: Columbia University; 2013-2020. Title: "Structure and Feedback in Cloud Service API Fuzzing." First position: Post-doc at Brown University.

Mathias Lécuyer: Columbia University, 2013–2019. Title: "Security, Privacy, and Transparency Guarantees for Machine Learning Systems." First positions: Microsoft Research Post-Doc (2019-2021) and Tenure-track Assistant Professor of Computer Science at University of British Columbia (starting in 2021).

Riley Spahn: Columbia University, 2013–2019. Title: "Data Protection in Emerging Mobile and Machine Learning Workloads." First position: Google.

**Graduated M.S. students who have co-authored research papers under my primary supervision:**

Mingen Pan: Columbia University, 2019–2020

Yannis Spiliopoulos: Columbia University, 2014–2016

Francis Lan: Columbia University, 2014–2015

Andrei Papancea: Columbia University, 2014–2015

Sravan Bhamidipati: Columbia University, 2012–2013

Nikhil Sarda: Columbia University, 2012–2013

Phillip Ames: Columbia University, 2012–2012

**Member of Ph.D. dissertation committees:**

Lingmei Weng, Ph.D. in Computer Science (Columbia University; defended April 2023; primary advisor: Jason Nieh).

Chengyu Lin, Ph.D. in Computer Science (Columbia University; defended October 2022; primary advisor: Tal Malkin).

Alex Van't Hof, Ph.D. in Computer Science (Columbia University; defended March 2022; primary advisor: Jason Nieh).

John Koh, Ph.D. in Computer Science (Columbia University; defended November 2020; primary advisors: Steve Bellovin, Jason Nieh).

Kevin Shi, Ph.D. in Computer Science (Columbia University; defended December 2019; primary advisor: Daniel Hsu).

Naser AlDuaij, Ph.D. in Computer Science (Columbia University; defended December 2019; primary advisor: Jason Nieh).

Yang Tang, Ph.D. in Computer Science (Columbia University; defended June 2019; primary advisor: Junfeng Yang).

Marios Ponomis, Ph.D. in Computer Science (Columbia University; defended August 2019; primary advisor: Angelos Keromytis).

Suphannee Sivakorn, Ph.D. in Computer Science (Columbia University; defended April 2018; primary advisor: Angelos Keromytis).

Christoffer Dall, Ph.D. in Computer Science (Columbia University; defended February 2017; primary advisor: Jason Nieh).

Yuanzhong Xu, Ph.D. in Computer Science (University of Texas at Austin; defended May 2016; primary advisor: Emmett Witchel).

Jonathan Bell, Ph.D. in Computer Science (Columbia University; defended April 2016; primary advisor: Gail Kaiser).

George Argyros, Ph.D. in Computer Science (Columbia University; defended March 2016; primary advisors: Angelos Keromytis and Tal Malkin).

Nicolas Viennot, Ph.D. in Computer Science (Columbia University; defended January 2016; primary advisor: Jason Nieh).

Yoshihisa Abe, Ph.D. in Computer Science (Carnegie Mellon University; defended December 2015; primary advisor Mahadev Satyanarayanan).

Vasileios Kemerlis, Ph.D. in Computer Science (Columbia University; defended August 2015; primary advisor: Angelos Keromytis).

Jeremy Andrus, Ph.D. in Computer Science (Columbia University; defended May 2015; primary advisor: Jason Nieh).

Kangkook Jee, Ph.D. in Computer Science (Columbia University; defended 2015; primary advisor: Angelos Keromytis).

Nathaniel Boggs, Ph.D. in Computer Science (Columbia University; defended 2015; primary advisor: Sal Stolfo).

Binh Vo, Ph.D. in Computer Science (Columbia University; defended 2015; primary advisor: Sal Stolfo).

Heming Cui, Ph.D. in Computer Science (Columbia University; defended 2014; primary advisor: Junfeng Yang).

Marcin Szczodrak, Ph.D. in Computer Science (Columbia University; defended 2014; primary advisor: Luca Carloni).

Jae Woo Lee, Ph.D. in Computer Science (Columbia University; defended 2013; primary advisor: Henning Schultzerinne).

# Service

## *Service to the University*

**Computer Science Department Service:**

Space committee, 2014-present.

Undergraduate advisor, 2023-present.

Ph.D. Fellowship committee, 2012-present.

Faculty Search committee, 2014, 2015, 2016.

**SEAS/University Service:**

Member of Columbia Data Science Institute (DSI) Cybersecurity Center, 2014-present.

Member of DSI's Systems for Data Science Center, 2018-present.

Facilitator for the Communications, Information, and Cybersecurity panel of the Strategic Discussion Forum, 2015.

Columbia IT facilities committee, 2014, 2015.

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

## *Conference, Workshop, and Journal Organization*

Program committee co-chair of the USENIX Conference on Operating Systems Design and Implementation (OSDI), 2023.

Associate editor of ACM Transactions on Computer Systems (TOCS), 2019-2020.

Program committee co-chair of the ACM Symposium on Cloud Computing (SoCC), 2019.

Privacy area organizer for Columbia's Trustworthy AI Conference, 2019.

Computer Science area co-chair, ACM Conference on Fairness, Accountability, and Transparency (FAT*), 2018, 2020.

Organizer of hands-on programming workshop for the Annual Engineering Exploration event organized by the Society of Women Engineers for New York City high-school female students, 2014, 2015, 2016.

Organizer (with Luis Ceze) of ISAT Workshop "The Future of Storage" on identifying break-through technologies on storage and data selectivity, 2016.

Organizer (with Katrina Ligett) of ISAT Workshop "Whither the Data" on understanding complex flows in data ecosystems, 2016.

Chair of Poster Session for USENIX Operating Systems Design and Implementation Conference (OSDI), 2014.

Chair of Poster Session for the European Conference on Computer Systems (EuroSys), 2013.

## *Conference Program Committees*

**Systems Conferences:**

USENIX Operating Systems Design and Implementation Conference (OSDI), 2024, 2022, 2020, 2018, 2016, 2014, 2012.

ACM Symposium on Operating Systems Principles (SOSP), 2021, 2017.

European Conference on Computer Systems (EuroSys), 2019, 2016, 2015, 2013.

USENIX Symposium on Networked Systems Design and Implementation (NSDI), 2019, 2018, 2011.

International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2016, 2020.

ACM Symposium on Cloud Computing (SoCC), 2015, 2013.

ACM Workshop on Hot Topics in Operating Systems (HotOS), 2019, 2017, 2013.

Hot Topics in Mobile Computing (HotMobile), 2016, 2013.

**Security and Privacy Conferences:**

IEEE Security and Privacy (S&P), 2017.

USENIX Security Symposium, 2015, 2013, 2012.

ACM Cloud Computing Security Workshop (CCSW), 2013.

Workshop on Data and Algorithmic Transparency (DAT), 2016.

**Outreach Conferences:**

Workshop on Technology and Consumer Protection (organized by the FTC), 2018, 2017.

Grace Hopper Celebration of Women in Computing Workshop (Grace Hopper), 2017.

## *Funding Proposal Reviewing*

NSF Panel for Formal Methods in the Field program, 2017.

NSF Panel for Secure and Trustworthy Computing program, 2014.

NSF Panel for Computer Systems Research, 2012.

## *Diversity Outreach Activities*

"Security and Privacy Guarantees in Machine Learning with Differential Privacy." Keynote speaker at the 8th Networking Networking Women (N2Women) workshop in conjunction with the ACM SenSys conference, 2019.

"Privacy in a Data-Driven World." Invited speaker at New York City Women in Data Science (WiDS) conference, 2018.

"Privacy in a Data-Driven World." Invited speaker at Columbia's NSF IGERT inter-disciplinary course "From Data to Solutions," 2018.

"Testing for Fairness." Invited panelist for the Negotiating Systemic Bias in Teaching event organized by Columbia University Libraries and the TOW Center for Journalism, 2018.

"Privacy in a Data-Driven World." Invited speaker at Columbia's NSF IGERT inter-disciplinary course "From Data to Solutions," 2017.

"Privacy in a Data-Driven World." Invited speaker at Columbia's Emerging Scholar Program, 2017.

"Privacy in a Data-Driven World." Invited speaker at Columbia's Engineering Women's Forum, 2017.

Organize hands-on programming workshop for the Annual Engineering Exploration event organized by the Society of Women Engineers for New York City high school female students, 2014, 2015, 2016. Participants are 23-25 high school students from all five boroughs of New York City.

Panelist at the 2016 Global Digital Futures Policy Forum, Columbia School of International and Public Affairs (SIPA). Invited to discuss the potential and pitfalls of an algorithmic society, 2016.

"Privacy in a Data-Driven World." Invited speaker at Columbia's NSF IGERT inter-disciplinary course "From Data to Solutions," 2016.

"Privacy in a Data-Driven World." Invited speaker at joint event organized by the Data Science Institute and Columbia Business School's Leadership and Ethics center, 2016.

"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.

"Toward a Transparent Web." Talk for Data Science Institute's industrial affiliates program, delivered for Bloomberg, 2014.

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

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

"Research and Education at Columbia's CS." Department host talk at Columbia's Engineering Women's Forum, 2012, 2013.

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

## Media Coverage

The data republic, "To safeguard democracy, the use of data should be made as transparent as possible," The Economist, 2016.

Priya Kumar, "When Was the Last Time You Read a Privacy Policy?" Slate.com, 2016.

Ben Johnson, Codebreaker episode #7 on National Public Radio (NPR) Tech Marketplace, 2016.

Tom Simonite, "Probing the Dark Side of Google's Ad-Targeting System," MIT Technology Review, 2015.

Jim Dwyer, "The Big Bang of Social Networking," The New York Times, 2014.

Steve Lohr, "XRay: A New Tool for Tracking the Use of Personal Data on the Web," The New York Times, 2014.

Leonard Lopate, "Two of 'The Brilliant 10' Scientists," The Leonard Lopate Show on WYNC, New York City's branch of NPR, 2014.

Bob Brown, "5 Cool Cloud Computing Research Projects," Network World, 2009.

John Markoff, "New Technology to Make Digital Data Self-Destruct," The New York Times, 2009.

Martin Kaste, "Digital Data Make For A Really Permanent Record," National Public Radio, The End of Privacy Series, 2009.

David Lee, "This Website Will Self-destruct...," BBC Digital Planet, 2009.

"This Message Will Self-destruct," The Economist, 2009.