Baishakhi Ray

ASSISTANT PROFESSOR · COLUMBIA UNIVERSITY

500 West 120 Street, Room 450, MC0401, New York City, New York, USA-10027

(+1) 303-748-2958 | rayb@cs.columbia.edu | http://rayb.info

Research Interest

I am primarily interested in Software Engineering research with a focus on improving software reliability and security. In particular, I devise novel program analysis techniques to analyze existing code properties and apply advanced machine learning and NLP models to learn from those properties. Such models help me building tools that automate program development, bug detection, and program repair for real-world large scale software.

Google Scholar ID: **https://scholar.google.com/citations?user=VaAEb5YAAAAJ&hl=en**

Experience

Columbia University — New York, New York, USA
ASSISTANT PROFESSOR — Jul 2018 - Current

University of Virginia — Charlottesville, Virginia, USA
ASSISTANT PROFESSOR — Oct 2015 - Jun 2018

University of California, Davis — Davis, California, USA
POSTDOCTORAL RESEARCH FELLOW — Oct 2013 - Sep 2015

The University of Texas at Austin — Austin, Texas, USA
GRADUATE RESEARCH ASSISTANT — Jan 2011 - May 2013

Education

Postdoc in Computer Science — Davis, California, USA
UNIVERSITY OF CALIFORNIA, DAVIS — 2013 - 2015
- Advisor: Prem Devanbu

PhD in Electrical & Computer Engineering — Austin, Texas, USA
UNIVERSITY OF TEXAS AT AUSTIN — 2010 - 2013
- GPA: 3.97
- Thesis: Analysis of Cross-System Porting and Porting Errors in Software Projects
- Advisor: Miryung Kim

MS in Computer Science — Boulder, Colorado, USA
UNIVERSITY OF COLORADO BOULDER — 2007 - 2009
- GPA: 4.0
- Thesis: SecureWear: Securing Wearable Mobile Social Network
- Advisor: Richard Han

B.Tech. in Computer Science — Kolkata, India
UNIVERSITY OF CALCUTTA — 2001 - 2004

B.Sc. in Physics — Kolkata, India
PRESIDENCY COLLEGE, KOLKATA — 1998 - 2001
- University Rank: 6

Honors & Awards

- 2017: Best Paper Award, SIGSOFT Symposium on the Foundations of Software Engineering (ESEC/FSE 2017)
- 2017: Best Paper Award, International Conference on Mining Software Repositories, (MSR 2017)
- 2017: CACM Research Highlight
- 2014: Best Practical Paper Award, IEEE Symposium on Security and Privacy (S&P Oakland), 2014
- 2013: Nominated for Distinguished Paper Award, Automated Software Engineering (ASE, 2013)
- 2012: Google Summer of Code
- 2001: Ranked 6 out of 15,000 students in B.Sc. Physics (Hons.) examination
- 2001: Jawaharlal Nehru Summer Scholarship for Advanced Research, India 2001

RESEARCH GRANTS

- 2018: NSF SPX: Integrating Persistent Memory in the Cloud
- 2018: NSF SaTC: CORE: EAGER: Finding Semantic Security Bugs with Pseudo-Oracle Testing
- 2016: NSF TWC: Small: Automated Detection and Repair of Error Handling Bugs in SSL/TLS Implementations
- 2016: NSF CHS: Small: Translating Compilers for Visual Computing in Dynamic Languages

Other Professional Experience

Microsoft Research   *Redmond, Washington, USA*
RESEARCH INTERN   *May 2013 - Aug 2013*
- Research Group: Empirical Software Engineering (ESE) in RiSE
- Mentors: Christian Bird, Nachiappan Nagappan, Thomas Zimmermann

Google Summer of Code, Google Inc.
RESEARCH INTERN   *May 2012 - Aug 2012*
- Mentors: Suzette Person, Neha Rungta, NASA

Avaya Research Lab   *Westminster, Colorado, USA*
RESEARCH INTERN   *May 2012 - Aug 2012*

Ericsson Pvt. Ltd.   *Boulder, Colorado, USA*
SOFTWARE ENGINEER   *Feb 2009 - Jun 2010*

Ixia, Sasken, and Texas Instruments   *India*
SOFTWARE ENGINEER   *Aug 2004 - Jun 2007*

Publications

CONFERENCE PUBLICATIONS [1]

2019

1. *NEUZZ: Efficient Fuzzing with Neural Program Smoothing.* D. She, K. Pei, D. Epstein, J. Yang, **B. Ray**, S. Jana. In 40th IEEE Symposium on Security and Privacy (**S&P'19**), May'19. 12 pages. acceptance rate: <20%.

2018

2. *DeepTest: Automated Testing of Deep-Neural-Network-driven Autonomous Cars.* Y. Tian, K. Pei, S. Jana, **B. Ray**. In 40th International Conference on Software Engineering (**ICSE'18**), May'18. 10 pages. acceptance rate: 20.9%.

3. *Building Language Models for Text with Named Entities.* Md R. Parvez, S. Chakraborty, **B. Ray**, K. Chang. In 56th Annual Meeting of the Association for Computational Linguistics (**ACL'18**), July'18. 10 pages. acceptance rate: 24%.

4. *Replay without Recording of Production Bugs for Service Oriented Applications.* N. Arora, J. Bell, F. Ivancic, G. Kaiser, **B. Ray**. In 33rd IEEE/ACM International Conference on Automated Software Engineering (**ASE'18**), Sep'18. 10 pages. acceptance rate: 19.9%.

5. *Evaluating How Developers Use General-Purpose Web-Search for Code Retrieval.* Md M. Rahman, J. Barson, S. Paul, J. Kayani, F. Lois, S. Quezada, C. Parnin, K. Stolee, and **B. Ray**. In 15th International Conference on Mining Software Repositories (MSR'18), May'18. 10 pages. acceptance rate: 33%.

6. *Obfuscation Resilient Search through Executable Classification.* F. Su, J. Bell, G. Kaiser, and **B. Ray**. In 2nd ACM SIGPLAN Workshop on Machine Learning and Programming Languages (MAPL'18), June'18. 10 pages.

2017

7. *Automatically Diagnosing and Repairing Error Handling Bugs in C.* Y. Tian, **B.Ray**. In 11th joint meeting of the European Software Engineering conference and the ACM Sigsoft Symposium on the Foundations of Software (**ESEC-FSE'17**), September'17. 10 pages. acceptance rate: 24.4%. ACM SIGSOFT Distinguished Paper Award.

8. *Some From Here, Some From There: Cross-Project Code Reuse in GitHub.* M. Gharehyazie, **B.Ray**, V. Filkov. In 14th International Conference on Mining Software Repositories (MSR'17), May'17. 10 pages. acceptance rate: 27%. ACM SIGSOFT Distinguished Paper Award.

9. *A Large Scale Study of Programming Languages and Code Quality in Github.* **B.Ray**, D. Posnett, P. T. Devanbu, V. Filkov. In Communications of the ACM, Oct'17. 8 pages. Invitation Only. CACM Research Highlights.

2016

10. *On the Naturalness of Buggy Code.* **B. Ray**, V. Hellendoorn, S. Godhane, Z. Tu, A. Bacchelli, P. Devanbu. In 38th International Conference on Software Engineering (**ICSE'16**), May'16. 10 pages. acceptance rate: 19%.

11. *Automatically Detecting Error Handling Bugs using Error Specifications.* S. Jana, Y. J. Kang, S. Roth, **B. Ray**. In 25th Usenix Security Symposium (**USENIX Security'16**), August'16. 18 pages. acceptance rate: 15.5%.

12. *APEx: Automated Inference of Error Specifications for C APIs.* Y. J. Kang, **B. Ray**, S. Jana. In 31st IEEE/ACM International Conference on Automated Software Engineering (**ASE'16**), November'16. 10 pages. acceptance rate: 19.1%.

2015

13. *Assert Use in GitHub Projects.* C. Casalnuovo, P. Devanbu, A. Oliveira, V. Filkov, **B. Ray**. In 37th International Conference on Software Engineering (**ICSE'15**), May'15 10 pages. acceptance rate: 18.5%.

14. *Gender and Tenure Diversity in GitHub Teams.* B. Vasilescu, D. Posnett, **B. Ray**, M. Brand, A. Serebrenik, P. Devanbu, V. Filkov. In International Conference on Human Factors in Computing Systems (**CHI'15**), April'15. 10 pages. acceptance rate: 23%.

---

[1] Top Tier conferences are marked in Blue. My students are highlighted with underline. My name is highlighted in **bold**.

2014

15. *A Large Scale Study of Programming Languages and Code Quality in Github.* **B. Ray**, D. Posnett, V. Filkov, P. T. Devanbu. In ACM SIGSOFT, 22nd International Symposium on the Foundations of Software Engineering (**FSE'14**), November'14. 10 pages. acceptance rate: 22%. Selected for publication in the "Research Highlights" section of the CACM. Media Coverage: SlashDot, The Register, Reddit, InfoWorld, Hacker News.

16. *Using Frankencerts for Automated Adversarial Testing of Certificate Validation in SSL/TLS Implementations.* C. Brubaker, S. Jana, **B. Ray**, S. Khurshid, and V. Shmatikov. In 35th IEEE Symposium on Security and Privacy, 2014 (**S&P Oakland'14**), May'14. 15 pages. acceptance rate: 13%. Best Practical Paper Award. Media Coverage: Reddit, Golem, Heise.

2013

17. *Detecting and Characterizing Semantic Inconsistencies in Ported Code.* **B. Ray**, M. Kim, S. Person, N. Rungta. In 28th IEEE/ACM International Conference on Automated Software Engineering, 2013 (**ASE'13**), November'13. 10 pages. acceptance rate: 23%. Nominated for Distinguished Paper Award, invited for ASE journal special issue.

18. *An Empirical Study of API Stability and Adoption in the Android Ecosystem.* T. McDonnell, **B. Ray**, M. Kim. In 29th IEEE International Conference on Software Maintenance, 2013 (ICSM'13), September'13. 10 pages. acceptance rate: 22%.

2012

19. *Case Study of Cross-System Porting in Forked Projects.* **B. Ray**, M. Kim. In ACM SIGSOFT, the 20th International Symposium on the Foundations of Software Engineering (**FSE'12**), November'12. 10 pages. acceptance rate: 17%.

20. *An Empirical Study of Supplementary Bug Fixes.* J. Park, M. Kim, **B. Ray**, D. Bae. In The 9th IEEE Working Conference on Mining Software Repositories (MSR'12), June'12. 10 pages. acceptance rate: 28%. Invited to the Special Issue of Journal of Empirical Software Engineering (EMSE).

2011

21. *PTask: Operating System Abstractions To Manage GPUs as Compute Devices.* CJ. Rossbach, J. Currey, M. Silberstein, **B. Ray**, E. Witchel. In Proceedings of the 23rd ACM Symposium on Operating System Principles (**SOSP'11**), October'11. 16 pages. acceptance rate: 17%.

Older

22. *Touch Me wE@r: Getting Physical with Social Networks.* A. Beach, **B. Ray**, L. Buechley. In 2009 Workshop on Sensor-based Models and Feedback Systems for Social Computing. Associated with SocialCom 2009. 6 pages.

23. *A Protocol for Building Secure and Reliable Covert Channel.* **B. Ray** and S. Mishra. In 6th Annual Conference on Privacy, Security and Trust, 2008. (PST'08). 8 pages.

24. *WhozThat? Evolving an Ecosystem for Context-Aware Mobile Social Networks.* A. Beach, **B. Ray**, et al. In IEEE Network Magazine Special Issue on Composable context aware services, 2008. 6 pages.

Short Papers

Tool Demo

25. *GitcProc: A Tool for Processing and Classifying GitHub Commits.* C. Casalnuovo, Y. Suchak, **B. Ray**, C. Rubio-Gonzalez. In International Symposium on Software Testing and Analysis (ISSTA'17 Tool-demo), July'17. 4 pages.

26. *Repertoire: A Cross-System Porting Analysis Tool for Forked Software Projects.* **B. Ray**, C. Wiley, M. Kim. In ACM SIGSOFT the 20th International Symposium on the Foundations of Software Engineering, Formal Research Tool Demonstration (FSE'12), November'12. 4 pages.

Poster

27. *Poster: Which Similarity Metric to Use for Software Documents? A Study on Information Retrieval-Based Software Engineering Tasks.* Md M. Rahman, S. Chakraborty, **B. Ray**. In 40th International Conference on Software Engineering (ICSE'18), May'18. 2 pages.

28. *Poster: Searching for High-performing Software Configurationswith Metaheuristic Algorithms.* C. Tang, K. Sullivan, **B. Ray**. In 40th International Conference on Software Engineering (ICSE'18), May'18. 2 pages.

29. *Poster: A Recommender System for Developer Onboarding* C. Liu, D. Yang, X. Zhang, H. Hu, J. Barson, **B. Ray**. In 40th International Conference on Software Engineering (ICSE'18), May'18. 2 pages.

Book Chapter

30. *A Large Ecosystem Study to Understand the Effect of Programming Languages on Code Quality.* **B. Ray**, D. Posnett. In Perspectives on Data Science for Software Engineering, Morgan Kaufmann, 2016.

31. *SecureWear: A Framework for Securing Mobile Social Networks.* **B. Ray**, R. Han. In Advances in Computer Science and Information Technology. Computer Science and Engineering, Vol. 85, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer Berlin Heidelberg, 2012. 10 pages.

Teaching

Undergraduate Course

- Programming Language & Translator: Columbia Universitya, Fall 2018
- Program Analysis for Robust Software: University of Virginia, Spring 2018

Graduate Course

- Empirical Software Engineering: University of Virginia, Spring 2017
- Software Engineering: Fall 2016
- Data Science In Software Engineering: University of Virginia, Spring 2016

Services

Program Co-Chair

- 2020:   FSE tool demo track.
- 2019:   ICSE Artifact Evaluation track
- 2019:   ASE: Late Breaking Result Track

Program Organizer

- 2019:   DeepTest Workshop (co-located at ICSE'19).
- 2019:   ML4SE Workshop (at Google Montreal).
- 2016:   Co-Organized NL+SE 2016 Workshop, co-located with FSE 2016.

Program Committee

- 2019:   ICSE Workshop, ICSE NIER, MSR, ASE, ICSME
- 2018:   ASE, FSE NIER, ISSTA Tool Demo, AsiaCCS, ICPC.
- 2017:   ASE (ERP), ICSE NIER, OOPSLA Onward, ICPC, MSR, ICSME.
- 2016:   ASE (ERP), MSR, ICSE Visions (V2025), ICSE Tool Demo, FSE SRC, NL+SE Workshop, ICSME Era, ISEC, APSEC.
- 2015:   MSR, MSR Challenge, ISEC.

Journal Review

- Transactions on Software Engineering (TSE).
- Empirical Software Engineering (EMSE).
- ACM Transactions on Software Engineering and Methodology (TOSEM).
- IEEE Transactions on Dependable and Secure Computing (TDSC).